

UNIVERSIDADE NOVE DE JULHO - UNINOVE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**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**

ANDRÉA MARTINIANO DA SILVA

SÃO PAULO
2012

ANDRÉA MARTINIANO DA SILVA

**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**

Dissertação apresentada ao Programa de
Mestrado em Engenharia de Produção da
Universidade Nove de Julho como parte dos
requisitos exigidos para a obtenção do grau de
Mestre em Engenharia de Produção

Orientador: Prof^o Dr. Renato José Sassi

**SÃO PAULO
2012**

Silva, Andrea Martiniano da.

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. / Andrea Martiniano da Silva. 2012.

130 f.

Dissertação (mestrado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2012.

Orientador (a): Prof. Dr. Renato José Sassi.

1. Absenteísmo. 2. Presenteísmo. 3. Mineração de Dados.

I. Sassi, Renato José. II. Título

CDU 624

**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**

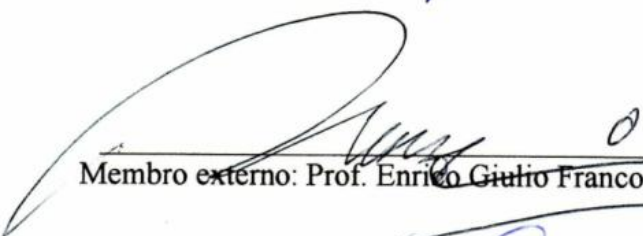
Por

ANDRÉA MARTINIANO DA SILVA

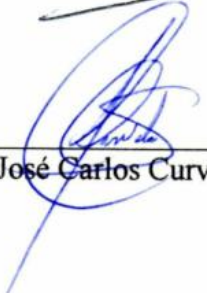
Dissertação apresentada ao Programa de Mestrado em Engenharia de Produção da Universidade Nove de Julho como parte dos requisitos exigidos para a obtenção do grau de Mestre em Engenharia de Produção, pela Banca Examinadora, formada por:



Presidente: Prof. Renato José Sassi, Dr. - Orientador, Uninove



Membro externo: Prof. Enrico Giulio Franco Polloni, Dr. - ISES



Membro interno: Prof. José Carlos Curvelo Santana, Dr. - Uninove

São Paulo, 28 de junho de 2012

Dedico este trabalho a minha querida Madrinha
Maria das Dores Corrêa (*in Memoriam*)

“Se Nossas mentes contêm processos que nos capacitam a solucionar problemas que consideramos difíceis. Então, inteligência é o nome que damos a qualquer um destes processos que ainda não compreendemos” Marvin Minsky

AGRADECIMENTOS

A Deus, o “Maravilhoso Conselheiro”, que nos ilumina, nos fortalece e nos consola. Devemos a Ele tudo o que temos recebido e tudo o que somos.

Àqueles que contribuíram, direta e indiretamente, para que este trabalho fosse realizado e, de maneira muito especial, a Ricardo Pinto Ferreira por estar presente em minha vida e o apoio em toda etapa deste trabalho.

Aos meus pais Jurandir Martiniano da Silva e Tereza Corrêa da Silva.

Ao Orientador Prof. Dr. Renato José Sassi, pelo incentivo e paciência em compartilhar seu conhecimento, que certamente contribuíram para meu aprendizado.

À Universidade Nove de Julho pela bolsa de estudos no Programa de Pós-Graduação em Engenharia de Produção.

Aos docentes e funcionários da Universidade Nove de Julho, em especial, ao Prof. Dr. André Felipe Henriques Librantz, ao Prof. Dr. Fábio Henrique Pereira.

Aos membros da banca, Prof. Dr. Enrico Giulio Franco Polloni, Prof. Dr. José Carlos Curvelo Santana, pelo pronto atendimento ao convite.

A todos os alunos do programa de mestrado em engenharia de produção, em especial, aos orientados do Prof. Dr. Renato José Sassi: Carlos Affonso, Walter Saraiva, Ricardo Pinto Ferreira, Adriano Arrivabene, Márcio Romero, Paulo Kaupa, Flávio Moreira da Costa, Tiago Vieira de Carvalho e Marcelo Drudi Miranda.

Aos alunos de Iniciação Científica e, de maneira especial, aos alunos Anselmo Bernardes Vieira e Rogério Pereira da Silva.

Ao Giuliano Lellis Ito Santos, pela revisão na língua vernácula.

Aos meus irmãos André Martiniano da Silva, Rogério Martiniano da Silva e Andreza Martiniano da Silva.

RESUMO

O desenvolvimento profissional, aliado ao das organizações e a pressão para atingir metas cada vez mais audaciosas, deixam os empregados sobrecarregados. Estes acabam adquirindo algum distúrbio no estado de saúde relacionado à atividade laboral, o que pode causar o absenteísmo ou o presenteísmo. Absenteísmo é o não comparecimento ao trabalho ou a obrigações sociais, conforme o programado. No caso do presenteísmo há a presença do empregado no trabalho ou a obrigações sociais, ainda que doente, contudo, suas atividades são improdutivas. Ambos são semelhantes no que diz respeito à perda de produtividade. Para a empresa isso significa que um serviço deixa de ser fornecido devido à falta do empregado ou é realizado de forma inadequada. De forma mais específica, para a empresa de Courier, em que a atividade principal é a entrega de encomendas expressas, o absenteísmo e o presenteísmo afetam a qualidade na prestação destes serviços. Para identificar estes fenômenos, este trabalho utilizou uma técnica da mineração de dados chamada rede neural artificial do tipo Mapa Auto-Organizável de Kohonen (*Self-Organizing Map*) ou simplesmente rede SOM para minerar os dados de uma base com 740 registros de 36 empregados de uma empresa de Courier da cidade de São Paulo. A rede SOM foi utilizada para agrupar os empregados pela semelhança dos seus atributos a fim de verificar o possível perfil absenteísta e ou presenteísta também presente neles e uma possível tendência a adquirirem esses perfis. Também foi realizada uma pesquisa do tipo (*survey*) com os mesmos 36 empregados com o objetivo de verificar a ocorrência de estresse nos empregados, motivo considerado fundamental para apresentar perfil absenteísta e ou presenteísta. Os resultados apresentados pela rede SOM mostraram que a aplicação da mineração de dados na identificação do perfil absenteísta e presenteísta em uma empresa de Courier na cidade de São Paulo é satisfatória, o que confirma a utilização desta técnica como uma boa opção a ser aplicada neste tipo de problema.

Palavras-chave: Absenteísmo, Presenteísmo, Mineração de Dados, Mapas Auto-organizáveis, Empresa de *Courier*.

ABSTRACT

Professional development, coupled with the organizations and the pressure to achieve goals more daring, leave employees overwhelmed. They end up getting some disturbance in health related to work activity, which can lead to absenteeism or presenteeism. Absenteeism is not reporting to work or social obligations, as scheduled. In the case of presenteeism is the presence of the employee at work or social obligations, though ill, however, their activities are unproductive. Both are similar with regard to loss of productivity. For the company this means that a service is no longer provided due to lack of employee or is performed improperly. More specifically, for the Courier company, where the main activity is the delivery of express shipments, absenteeism and presenteeism affect the quality of these services. To identify these phenomena, this study used a data mining technique called artificial neural network-type Self-Organizing Map of Kohonen (Self-Organizing Map) or simply SOM to mine data from a database with 740 records of 36 employees Courier firm the city of Sao Paulo. The SOM was used to group employees by similarity of their attributes in order to verify the possible profile and absenteeist or presenteeist also present in them and a possible tendency to acquire these profiles. Also a survey was conducted with the same the with 36 employees to verify the occurrence of stress in employees, and therefore considered essential to present profile absenteeist and presenteeist. The results presented by SOM showed that application of data mining to identify and profile absenteeist and presenteeist in a Courier company in São Paulo is satisfactory, which confirms the use of this technique as a good option to be applied in such problem.

Keywords: Absenteeism, Presenteeism, Data Mining, Self-organizing Map, Courier Company.

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	ABSENTEÍSMO.....	17
1.2	PRESENTEÍSMO.....	18
1.2.1	O Absenteísmo e o Presenteísmo.....	19
1.3	MINERAÇÃO DE DADOS.....	19
1.4	REDES NEURAIIS ARTIFICIAIS.....	20
1.4.1	O Mapa Auto-Organizável de Kohonen.....	20
1.4.2	A <i>Multilayer Perceptron</i>	21
1.5	MOTIVAÇÃO.....	21
1.6	JUSTIFICATIVA.....	22
1.7	OBJETIVOS.....	23
1.7.1	Objetivo Geral.....	23
1.7.2	Objetivos Específicos.....	23
1.8	METODOLOGIA.....	23
1.9	ORGANIZAÇÃO DO TRABALHO.....	24
2	FUNDAMENTAÇÃO TEÓRICA.....	25
2.1	O ABSENTEÍSMO.....	25
2.1.1	Um breve histórico do Absenteísmo.....	26
2.1.2	O Absenteísmo nas Empresas.....	27
2.1.3	As Causas do Absenteísmo.....	28
2.1.4	Consequências do Absenteísmo nas Empresas.....	31
2.1.5	O Custo do Absenteísmo.....	32
2.1.6	Quantificação do Índice de Absenteísmo.....	33
2.2	O PRESENTEÍSMO.....	35

2.2.1	O presenteísmo nas Empresas.....	37
2.2.2	As causas do presenteísmo.....	38
2.2.3	O custo do presenteísmo.....	40
2.2.4	A influência do presenteísmo absenteísmo.....	41
2.2.5	Quantificação do índice de presenteísmo.....	43
2.3	DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS.....	44
2.3.1	Fases do KDD.....	45
2.3.2	Tarefas do KDD.....	51
2.4	O MAPA AUTO-ORGANIZÁVEL DE KOHONEN.....	53
2.4.1	características da rede SOM.....	55
2.4.2	O algoritmo de aprendizado da rede SOM.....	56
2.4.3	O processo competitivo da rede SOM.....	57
2.4.4	O processo cooperativo da rede SOM.....	61
2.4.5	O processo adaptativo sináptico da rede SOM.....	64
2.4.6	As duas fases do processo adaptativo da rede SOM.....	66
2.4.7	Medidas de qualidade da rede SOM.....	66
2.5	<i>MULTILAYER PERCEPTRON</i>	67
2.5.1	Critérios de parada de treinamento da MLP.....	69
3	METODOLOGIA DA PESQUISA.....	71
3.1.	BASE DE DADOS EXPERIMENTAL.....	71
3.1.1	REALIZAÇÃO DOS EXPERIMENTOS.....	76
3.2	FERRAMENTAS UTILIZADAS E PLATAFORMAS DE ENSAIO.....	76
3.3	PARÂMETROS DA REDE SOM.....	77
3.4	PARÂMETROS DA <i>MULTILAYER PERCEPTRON</i>	78
3.5	CARACTERÍSTICAS DA PESQUISA DO TIPO <i>SURVEY</i>	78

4	RESULTADOS E DISCUSSÕES.....	80
4.1	RESULTADOS DA PESQUISA DO TIPO <i>SURVEY</i>	80
4.2	REALIZAÇÃO DOS EXPERIMENTOS COM A REDE SOM.....	95
4.2.1	REALIZAÇÃO DOS EXPERIMENTOS COM A REDE MLP.....	106
4.3	DISCUSSÃO DOS RESULTADOS.....	109
5	CONCLUSÃO	113
	REFERÊNCIAS BIBLIOGRÁFICAS.....	115
	APÊNDICE A – QUESTIONÁRIO.....	126
	APÊNDICE B – PRINCIPAIS PUBLICAÇÕES DO AUTOR.....	127

LISTA DE FIGURAS

Figura 2.1 – Países mais estressados do mundo.....	39
Figura 2.2 – Estimativa sobre os custos do presenteísmo e as condições específicas de saúde.....	41
Figura 2.3 – Custos com a saúde ocupacional.....	42
Figura 2.4 – Fases do processo de KDD.....	46
Figura 2.5 – Esforço requerido para cada fase do processo de KDD.....	46
Figura 2.6 – Esquema simplificado de mineração de dados.....	50
Figura 2.7– Representação do neurônio biológico e artificial.....	53
Figura 2.8 – Mapa com arranjo hexagonal para uma rede SOM com grade bidimensional.....	56
Figura 2.9 – Modelo da rede SOM.....	58
Figura 2.10 – Processo de ajuste do vetor de pesos do neurônio vencedor.....	59
Figura 2.11 – Distribuição em espaço bidimensional dos vetores.....	60
Figura 2.12 – Função de vizinhança gaussiana.....	62
Figura 2.13 – Vizinhança topológica.....	62
Figura 2.14 – Estrutura básica da RNA do tipo MLP.....	67
Figura 2.15 – Função sigmóide.....	69
Figura 3.1 – Metodologia experimental.....	76
Figura 4.1 – <i>Clusters</i> (Questão 1).....	81
Figura 4.2 – Distribuição relativa (Questão 1).....	81
Figura 4.3 – <i>Clusters</i> (Questão 2).....	82
Figura 4.4 – Distribuição relativa (Questão 2).....	82
Figura 4.5 – <i>Clusters</i> (Questão 3).....	83
Figura 4.6 – Distribuição relativa (Questão 3).....	83
Figura 4.7 – <i>Clusters</i> (Questão 4).....	84
Figura 4.8 – Distribuição relativa (Questão 4).....	84
Figura 4.9 – <i>Clusters</i> (Questão 5).....	85
Figura 4.10 – Distribuição relativa (Questão 5).....	85
Figura 4.11 – <i>Clusters</i> (Questão 6).....	86
Figura 4.12 – Distribuição relativa (Questão 6).....	86
Figura 4.13 – <i>Clusters</i> (Questão 7).....	87
Figura 4.14 – Distribuição relativa (Questão 7).....	87
Figura 4.15 – <i>Clusters</i> (Questão 8).....	88
Figura 4.16 – Distribuição relativa (Questão 8).....	88

Figura 4.17 – <i>Clusters</i> (Questão 9).....	89
Figura 4.18 – Distribuição relativa (Questão 9).....	89
Figura 4.19 – <i>Clusters</i> (Questão 10).....	90
Figura 4.20 – Distribuição relativa (Questão 10).....	90
Figura 4.21 – <i>Clusters</i> (Questão 11).....	91
Figura 4.22 – Distribuição relativa (Questão 11).....	91
Figura 4.23 – <i>Clusters</i> (Questão 12).....	92
Figura 4.24 – Distribuição relativa (Questão 12).....	92
Figura 4.25 – <i>Clusters</i> (Questão 13).....	93
Figura 4.26 – Distribuição relativa (Questão 13).....	93
Figura 4.27 – <i>Clusters</i> (Questão 14).....	94
Figura 4.28 – Distribuição relativa (Questão 14).....	94
Figura 4.29 Rede SOM rotulada.....	95
Figura 4.30 – Atributo Mês.....	96
Figura 4.31 – Atributo dia da semana.....	98
Figura 4.32 – Atributo estação do ano.....	99
Figura 4.33 – Atributo tempo de afastamento em horas.....	100
Figura 4.34 – Atributo distância da residência para o trabalho em quilômetros.....	101
Figura 4.35 – Atributo tempo de serviço.....	102
Figura 4.36 – Atributo idade.....	103
Figura 4.37 – Atributo escolaridade.....	104
Figura 4.38 – Atributo número de filhos.....	105
Figura 4.39 – Atributo índice de massa corporal.....	106
Figura 4.40 – Resultado da MLP na classificação da rede SOM rotulada.....	108

LISTA DE TABELAS

Tabela 3.1 – Lista de categorias com (CID).....	73
Tabela 3.2 – Lista de categorias sem (CID).....	74
Tabela 3.3 – Base de dados do absenteísmo.....	75
Tabela 4.1 – Absenteísmo em horas.....	96
Tabela 4.2 – Meses <i>versus</i> absenteísmo.....	97
Tabela 4.3 – Dia da semana <i>versus</i> absenteísmo.....	98
Tabela 4.4 – Estação do ano <i>versus</i> absenteísmo.....	99
Tabela 4.5 – Tempo de afastamento em horas.....	100
Tabela 4.6 – Distância da residência para o trabalho em quilômetros.....	101
Tabela 4.7 – Tempo de serviço.....	102
Tabela 4.8 – Idade <i>versus</i> absenteísmo.....	103
Tabela 4.9 – Escolaridade <i>versus</i> absenteísmo.....	104
Tabela 4.10 – Número de filhos <i>versus</i> absenteísmo.....	105
Tabela 4.11 – Índice de massa corporal <i>versus</i> absenteísmo.....	106
Tabela 4.12 – Sumário dos perfis estudados.....	109

LISTA DE ABREVIATURAS E SIGLAS

BMU	-	<i>Best Match Unit</i>
CID	-	Código Internacional de Doenças
DCBD	-	Descoberta de Conhecimento em Base de Dados
DM	-	<i>Data Mining</i>
DOU	-	Diário Oficial da União
GB	-	<i>GigaByte</i>
IA	-	Inteligência Artificial
KDD	-	<i>Knowledge Discovery in Databases</i>
MLP	-	<i>MultiLayer Perceptron</i>
OMS	-	Organização Mundial de Saúde
QVT	-	Qualidade de Vida no Trabalho
RAM	-	<i>Random Access Memory</i>
RNA	-	Rede Neural Artificial
RNAs	-	Redes Neurais Artificiais
SOM	-	<i>Self-Organizing Maps</i>

1 INTRODUÇÃO

A evolução da sociedade ressaltou a importância do relacionamento entre pessoas, diferentes culturas e mercados. As nações e as empresas vivenciam constantes avanços tecnológico, sócio-cultural 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 e 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.

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.

O absenteísmo em geral é definido como o não comparecimento do empregado no trabalho, conforme o programado (JOHNS, 2010). 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 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.

As empresas se utilizam de diversas técnicas para tentar detectar e prevenir o absenteísmo e o presenteísmo entre os seus empregados (THORWARTH *et al.*, 2009). Assim, de acordo com Cascio e Boudreau (2008) problemas podem ser evitados ao reconhecer que o empregado pode não estar disponível para realizar seu trabalho, conforme programado.

Para as empresas, que querem permanecer no mercado competitivo, é de suma importância um planejamento voltado para o combate dos fatores de absenteísmo e de presenteísmo. Haja vista, a relação evidente entre esses fatores e a baixa qualidade dos

serviços prestados nos diferentes ramos de produção (DALL'INHA, 2006).

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, que podem ser definidos como o processo de extração de conhecimentos válidos, novos, potencialmente úteis e compreensíveis para apoiar a tomada de decisão (FAYYAD *et al.*, 1996a).

O KDD é um processo interativo e iterativo formado por diversas fases, sendo a principal o *data mining* ou a mineração de dados. A mineração de dados é o elemento responsável pela extração do conhecimento implícito e útil contido em um banco de dados, utilizando, para isso, por exemplo, uma técnica da Inteligência Artificial (IA) como a rede neural artificial.

1.1 ABSENTEÍSMO

É conhecido também por absentismo ou ausentismo, termo que designa a falta do empregado em seu local de trabalho, mesmo não sendo motivada por doença prolongada ou licença legal (CHIAVENATO, 1994).

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 (GEHRING JUNIOR, 2007; PENATTI *et al.*, 2006).

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.

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 *et al.*, 2006).

1.2 PRESENTEÍSMO

O presenteísmo começou a ser estudado na década de 50, na França, e tornou-se mais conhecido durante os últimos anos do século XX. No Brasil o fenômeno é pouco conhecido e os departamentos de Recursos Humanos ainda não o veem como uma ameaça a produtividade e a lucratividade da empresa (GOETZEL *et al.*, 2004).

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.

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. 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).

1.2.1 O Absenteísmo e o 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.

Por outro lado, 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 *et al.*, 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 *et al.*, 1997).

1.3 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.

A mineração de dados permite a tomada de decisão baseada em fatos e não em suposições, podendo responder a questões que tradicionalmente demandariam muito tempo para resolver. Ela 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.*, 1996a; CABENA *et al.*, 1998).

Uma das grandes vantagens de se utilizar as técnicas de mineração de dados, 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 em inteligência artificial, 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).

1.4 REDES NEURAIAS ARTIFICIAIS

As Redes Neurais Artificiais (RNAs) são modelos constituídos por unidades de processamento simples, chamados de neurônios artificiais, que calculam funções matemáticas. Estes modelos são inspirados na estrutura do cérebro e têm como objetivo simular o comportamento humano, tais como: aprendizagem, associação, generalização e abstração quando submetidas a treinamento.

As RNAs são particularmente eficientes para o mapeamento entrada/saída de sistemas não lineares e para realizar processamento paralelo, além de simular sistemas complexos (HAYKIN, 2001). As RNAs generalizam os resultados obtidos para dados previamente desconhecidos, ou seja, produzem respostas coerentes e apropriadas para padrões ou exemplos que não foram utilizados no seu treinamento (HAYKIN, 2001). Para Kovács (2006), as RNAs são consideradas como um aproximador universal de funções contínuas.

As RNAs podem ser aplicadas em diversas áreas e na solução de vários problemas como: medicina (BLAZADONAKIS e MICHALIS, 2008), previsão de risco de crédito (SELAU e RIBEIRO, 2009), indústria automotiva (AFFONSO, 2010), previsão do comportamento do tráfego veicular urbano (FERREIRA, 2011), *data mining* (BIGUS, 1996), reconhecimento de padrões (HAYKIN, 2001) e em associação com outras técnicas (SASSI, 2012).

1.4.1 O Mapa Auto-Organizável de Kohonen

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).

Em vários trabalhos, esta rede é chamada de rede de Kohonen, o que pode provocar certa confusão, uma vez que Kohonen também propôs outros modelos de RNAs (BRAGA *et al.*, 2011). Neste trabalho serão chamadas simplesmente de rede SOM.

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.

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).

1.4.2 A *Multilayer Perceptron*

Uma RNA do tipo *Multilayer Perceptron* (MLP) ou *Perceptron* de Múltiplas Camadas é constituída de no mínimo três camadas de neurônios, sendo uma camada de entrada, uma de saída e no mínimo uma camada intermediária (ou escondida). Os sinais de entrada são propagados de uma camada para a seguinte até atingir a saída da rede.

A rede MLP normalmente é aplicada em problemas de classificação (CARVALHO, 2005), de aproximação (ou análise de regressão) que inclui previsão e modelagem de séries temporais (HAYKIN, 2001) em áreas como: controle (MITCHELL, 1997), diagnóstico (SILVA, 2009) e *data mining* (BIGUS, 1996).

1.5 MOTIVAÇÃO

A pesquisa sobre o absentéismo e o presenteísmo foi motivada pelo desafio de identificar o perfil de empregados absentéistas e presenteístas em uma empresa de *Courier* da cidade de São Paulo, e empregados que podem apresentar os dois comportamentos.

De acordo com Cunha *et al.* (2009), a pesquisa com o setor público, saúde do trabalhador, trabalhadores, perfil de saúde, absentéismo e licença médica apontou escassos estudos sobre absentéismo e restringiram-se a alguns grupos de profissionais, tais como servidores de hospitais e de universidades.

O estudo de Cunha *et al.* (2009) mostra ainda que, em se tratando de servidores públicos, essa lacuna é ainda maior, uma vez que são raros os dados consolidados sobre a saúde desse grupo de empregados no país.

Dessa forma, conclui-se que a pesquisa e a aplicação de novas técnicas para identificar tais comportamentos em profissionais que não fazem parte dos servidores de hospitais e universidades são decisivas para expansão das pesquisas sobre o absentéismo e o presenteísmo.

1.6 JUSTIFICATIVA

Levin-Epstein (2005) afirma que a perda de produtividade devido ao presenteísmo ocorre quando os empregados estão no trabalho, porém com rendimento reduzido devido a algum tipo de doença. Enquanto o absenteísmo tem sido muito pesquisado, o mesmo não ocorre com o presenteísmo.

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 *et al.*, 2006; CUNHA *et al.*, 2009).

Em relação ao presenteísmo de acordo com Hemp (2004), o que pode parecer controverso, é que o presenteísmo pode custar substancialmente mais do que os custos diretos nos tratamentos médicos e farmacêuticos, cerca de 20% a mais. Já os custos médicos indiretos com a saúde estão relacionados com o presenteísmo e representam cerca de 60% a mais.

Por possuírem grandes quantidades de informações, as empresas cada vez mais necessitam de ajuda de técnicas computacionais apropriadas para auxiliar na análise, na interpretação e no relacionamento dessas informações em busca de conhecimento.

Assim, técnicas de Inteligência Artificial, como as RNAs surgem como alternativa de apoio ou substituição dos métodos tradicionais de processamento da informação para ajudar a minerar, extrair informações valiosas de uma base de dados na determinação do perfil dos empregados que possam apresentar tendência ao absenteísmo e ao presenteísmo.

Como justificativa principal da pesquisa, é necessário citar o desafio de identificar o perfil absenteísta e presenteísta, que impulsiona a utilização da mineração de dados, buscando a obtenção de resultados que possibilitem uma análise mais aprofundada e a consequente identificação do perfil do empregado que apresenta um dos dois fenômenos.

O estudo em uma empresa de *Courier* da cidade de São Paulo deve-se ainda a preocupação da empresa com o absenteísmo e com o presenteísmo, tais constatações foram observadas através de documentos internos que discorrem sobre o assunto, podem-se citar: relatórios operacionais, gerenciais e estratégicos da empresa.

1.7 OBJETIVOS

1.7.1 Objetivo Geral

O objetivo geral deste trabalho é aplicar técnica de mineração de dados para identificar o perfil de empregados absenteístas e/ou presenteístas em empresa de *Courier* da Cidade de São Paulo.

1.7.2 Objetivos Específicos

Realizar uma pesquisa tipo *survey* para verificar o nível de estresse percebido pelos empregados.

- Aquisição da base de dados.

- Desenvolver uma rede SOM e aplicá-la na base de dados dos empregados da empresa de *courier* para que sejam gerados os agrupamentos.

- Desenvolver uma rede MLP e aplicá-la para classificar o perfil dos empregados agrupados pela rede SOM. Desta forma, verifica-se se a rede SOM agrupou os empregados adequadamente, ou seja, os agrupamentos são confiáveis para determinação do perfil dos empregados.

1.8 METODOLOGIA

A metodologia adotada foi definida como bibliográfica, exploratória e experimental.

A realização da pesquisa está embasada em consultas às fontes bibliográficas e de referencial teóricos como: artigos, livros, teses, dissertações e *websites* com conteúdos sobre absenteísmo, presenteísmo, mineração de dados, RNAs, rede SOM e a rede MLP.

Foi desenvolvida e aplicada uma rede SOM como técnica de mineração de dados para agrupar os empregados que apresentam o perfil absenteísta e/ou presenteísta. Também foi desenvolvida e aplicada uma rede MLP para, por meio da classificação, validar os agrupamentos gerados pela rede SOM.

Adicionalmente foi realizada uma pesquisa do tipo *survey* com os mesmos empregados. A pesquisa teve como objetivo verificar o estresse, percebido pelos empregados, e as possíveis relações com o absenteísmo e o presenteísmo.

1.9 ORGANIZAÇÃO DO TRABALHO

Este trabalho está estruturado em quatro capítulos, além deste capítulo introdutório, o trabalho compõe-se das seguintes partes:

Capítulo 2 – Fundamentação Teórica. Neste capítulo é apresentada a fundamentação teórica do Absenteísmo, do Presenteísmo, da Descoberta de Conhecimento em Base de Dados e dos Mapas Auto-Organizáveis de Kohonen.

Capítulo 3 – Metodologia da Pesquisa. Neste capítulo é apresentada a metodologia da pesquisa.

Capítulo 4 – Realização de Experimentos e Discussão de Resultados. Neste capítulo são apresentados os resultados da *survey* e dos experimentos com a base de dados realizada com a rede SOM e MLP. Finalizando o capítulo com a discussão dos resultados.

Capítulo 5 – Neste capítulo apresentam-se as Conclusões.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo é apresentada a fundamentação teórica sobre o absenteísmo, o presenteísmo, a descoberta de conhecimento em base de dados, o mapa auto-organizável de Kohonen e a rede *Multilayer Perceptron* (MLP).

2.1 O ABSENTEÍSMO

Para acompanhar os impactos da globalização, os avanços tecnológicos produzem mudanças nas organizações empresariais, além da sociedade como um todo, exigindo alterações na forma de agir diante desta nova realidade (McGREE e PRUSAK, 1994).

Foina (2006) destacou que, com a globalização e o rápido desenvolvimento das empresas, os mercados tendem a crescer e ampliar ainda mais suas atividades, o que é inevitável. Em função desse desenvolvimento a pressão por atingir metas torna-se cada vez mais audaciosas.

Neste aspecto, o trabalho voltado para o combate dos fatores absenteísmo e presenteísmo é de grande importância para as empresas, que querem permanecer autossustentáveis no mercado, haja vista a relação evidente entre esses fatores e a baixa qualidade dos serviços prestados nos diferentes ramos de produção.

No entanto, as empresas estão cada vez mais preocupadas com o bem estar de seus empregados, pois investem e implantam tecnologias direcionadas para a Qualidade de Vida no Trabalho (QVT). A QVT relaciona-se com o comprometimento pessoal e a participação com o bem-estar do empregado na execução da tarefa na empresa. Afinal um ambiente organizacional onde há uma gestão dinâmica e contingencial de fatores físicos, sociológicos, psicológicos e tecnológicos da empresa do próprio trabalho torna-se saudável e mais propício ao aumento de produtividade (FREITAS e SOUZA, 2009).

Martel e Dupuis (2006) destacaram que a QVT é fortemente influenciada pelo cenário econômico de um determinado período. Esta realidade é caracterizada por diversos aspectos, dentre os quais:

- globalização do mercado, resultando em crescente competição entre nações e, por conseguinte, entre empresas;
- a emergência de países onde os custos de produção são menores; e

- o crescimento de tecnologias de comunicação.

É notória a importância dada às questões relacionadas à QVT. De acordo com Lacanna (2008) a QVT é uma poderosa arma de competitividade do mercado, além de melhorar o desempenho e a produtividade, aumenta a confiabilidade do cliente.

2.1.1 Um breve histórico do Absenteísmo

O termo originou-se da palavra absentismo aplicado aos proprietários rurais que abandonavam o campo para viver na cidade. No período industrial do século XVIII, esse termo foi aplicado aos trabalhadores que faltavam ao trabalho e apresentava características incompatíveis com o bem-estar do empregado, percebe-se, deste modo, tratar-se de um problema antigo que até hoje gera transtornos às empresas (PENATTI *et al.*, 2006; AGUIAR e OLIVEIRA, 2009).

O absenteísmo, também conhecido como absentismo e ausentismo, é conceituado de diversas formas de acordo com a concepção de diferentes autores. Em Chiavenato (1994), o absenteísmo é considerado a falta ou a ausência de um empregado em seu ambiente de trabalho e pode ser definido também como incapacidade temporária ou permanente dessa ausência. Segundo Gehring Junior *et al.* (2007) o termo absenteísmo, significa a falta de assiduidade ao trabalho ou a outras obrigações sociais. Johns (2010) vê o absenteísmo como o não comparecimento no trabalho conforme o programado. Isso geralmente significa que o trabalho é feito de forma menos eficiente por outro empregado ou não é feito.

Apresenta-se a seguir as cinco categorias do absenteísmo (INOUE *et al.*, 2008):

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

Em relação à organização do trabalho, o absenteísmo tem se traduzido em insatisfação, desmotivação e sobrecarga da equipe de trabalho e pode ser indicativo da existência de problemas extremamente preocupantes quando ocasionado por doença (SILVA e MARZIALE, 2012a).

Para Alves *et al.* (2006), a conciliação da expectativa do empregado com as necessidades organizacionais tornou-se um desafio diante das dificuldades apresentadas pelo ambiente de trabalho como situações de insalubridade, longas jornadas de trabalho, baixa remuneração, duplo emprego, tensão emocional, que podem ter como uma das consequências, a falta ao trabalho ou absenteísmo.

Para Bachouch *et al.* (2010) as atribuições de atividades aos empregados é uma tarefa difícil. Erros cometidos em tais atribuições podem ter consequências em grandes proporções, como a eficiência reduzida, a falta de satisfação no trabalho e até mesmo o absentismo.

O tipo de atividade profissional e as condições em que a mesma é desenvolvida podem determinar os riscos para a saúde do trabalhador. Estão relacionados ao ambiente de trabalho fatores como ruído, iluminação inadequada, temperaturas extremas, vibração e higiene.

Relacionados à organização do trabalho estão fatores como conteúdo da tarefa, horário de trabalho, sobrecarga ou subcarga de trabalho físico e mental, participação e envolvimento no trabalho, relações interpessoais no trabalho, ritmo de trabalho e pressão do tempo, impossibilidade de ascensão funcional, pelas ausências de implementação de planos de cargos e salários (GARCIA e SILVA, 2009).

Segundo Inoue *et al.* (2008) para que as empresas consigam identificar o real motivo das ausências dos seus empregados, faz-se necessário um estudo mais detalhado das causas e consequências do absenteísmo.

2.1.2 O Absenteísmo nas Empresas

Cada empresa considera o absenteísmo de uma forma diferente, de acordo com sua realidade organizacional, bem como a realidade de seus empregados. Ou seja, em uma determinada empresa é possível que os gestores considerem como absenteísmo a falta injustificada e imprevista do empregado, já em outra empresa, conforme afirmam Silva e Marziale (2012b), pode-se considerar o absenteísmo como qualquer ausência do trabalhador, incluindo os afastamentos legais (e por isso previsto), que são de direito do empregado, tais como licença maternidade, férias, abonos, etc.

A cultura relativa ao absenteísmo na empresa é um dos assuntos mais pesquisados, pois se entende que o trabalhador pode achar normal certo número de faltas anuais, como, por exemplo, uma vez que seus colegas se comportem da mesma forma e não exista, por parte da empresa, qualquer restrição a esse padrão (DIAS, 2005).

A observação de fatores como condições de trabalho, natureza da supervisão, estilos de liderança, participação na tomada de decisões e relacionamentos profissionais pode ajudar a entender o impacto do absenteísmo nas empresas (PENATTI *et al.*, 2006).

Freitas e Souza (2009) destacaram que num ambiente organizacional onde há uma gestão dinâmica e contingencial de fatores físicos, sociológicos, psicológicos e tecnológicos da organização do próprio trabalho, têm-se um ambiente saudável e mais propício ao aumento de produtividade.

A importância das necessidades humanas varia conforme a cultura de cada indivíduo e de cada empresa Chiavenato (1999), daí que a QVT não é determinada apenas pelas características individuais e nem pelas características organizacionais, mas sim pela interação sistêmica entre ambas.

2.1.3 As Causas do Absenteísmo

A saúde do trabalho, em nível mundial, é considerada um pilar fundamental para o desenvolvimento de um país e para suas ações dirigidas à promoção e à proteção dos empregados.

A classe trabalhadora desempenha importante papel na economia, independentemente das condições socioeconômicas encontradas, pois dela depende a parcela produtiva da indústria. A falta ao trabalho por motivo de doença, ou absenteísmo doença, acarreta o rompimento ou o deslocamento do equilíbrio formado entre a saúde do empregado e a produtividade (DIACOV e LIMA, 1988).

À medida que a sociedade avança econômica, social e tecnologicamente, aumenta a necessidade de ações efetivas com relação à saúde do empregado (DIACOV e LIMA, 1988). Esta mesma necessidade de avanço faz com que o gerenciamento do absenteísmo exija muita atenção das empresas e gere muitas preocupações ligadas a diversos fatores, que vão desde questões sociais, de saúde, de gestão de pessoas, dentre outros problemas como excesso de trabalho, tornando, assim, este tema complexo e difícil de ser gerenciado (PENATTI *et al.*, 2006).

Hito (2007) enumerou as principais causas do absenteísmo: doença efetivamente comprovada e não comprovada, razões diversas de caráter familiar, atrasos involuntários ou por motivos de força maior, faltas voluntárias por motivos pessoais, dificuldades e problemas financeiros, problemas de transporte, baixa motivação para trabalhar, supervisão precária da chefia, políticas inadequadas da empresa, jornadas prolongadas de trabalho, condições insalubres e a tensão emocional.

Gehring Junior *et al.* (2007) relataram que a ausência no trabalho por profissionais de enfermagem da rede básica do SUS em Campinas era influenciada por inúmeros fatores, intrínsecas ou extrínsecas ao trabalho, que atuam sobre a pessoa tornando o absenteísmo um problema de alta complexidade, este fenômeno pode manifestar-se de diversas maneiras: desorganização do trabalho em equipe, sobrecarga de trabalho, insatisfação dos empregados, queda na qualidade e quantidade de trabalho realizado. Dessa forma são afetados os interesses da empresa e os direitos do empregado, bem como a relação de trabalho entre ambos.

Há uma relação direta entre o tipo de trabalho e as causas de afastamento entre os empregados com maior perda no absenteísmo de média duração (03 a 15 dias de afastamento) e o maior número de ocorrências no absenteísmo de curta duração (01 a 02 dias de afastamento), associando o afastamento do trabalho a fatores sociais, culturais e do trabalho (HITO, 2007).

Em termos práticos, o problema do absenteísmo é inversamente proporcional à satisfação no trabalho, Penatti *et al.* (2006).

O absenteísmo pode ser atribuído a causas conhecidas e a causas ignoradas ou previstas e não-previstas. Dentre as conhecidas (previstas), estão todas as amparadas por lei e que estão inseridas no estatuto do trabalhador ou no acordo coletivo de trabalho, por isso mesmo, justificada ao empregador, solicitando-lhe a permissão de ausência com antecedência: férias, folgas, feriados e outras.

As causas ignoradas (não-previstas), são as que efetivamente, caracterizam o absenteísmo, pois seu caráter é imprevisível e inevitável, são justificadas geralmente por problemas de saúde do empregado e/ou de seus dependentes, além de fatores aleatórios dos mais diversos como as faltas abonadas e as faltas injustificadas, licença maternidade e paternidade, adoção, acidente de trabalho e outras licenças amparadas por Lei (PENATTI *et al.*, 2006; MALLADA, 2012).

De acordo com NG *et al.* (2008), o absenteísmo imprevisto incorre em perdas tangíveis e intangíveis que se manifestam em prazos não cumpridos, custos de horas extras, o ânimo dos empregados diminui e a eficiência é reduzida.

Porém, Chiavenato (1999) declarou que as causas do absenteísmo podem decorrer também da organização do trabalho, da supervisão deficiente, de tarefas sem muito estímulo, da falta de motivação da equipe e do estilo de gerenciar da empresa, o absenteísmo não é apenas um indicador de adoecimento do empregado, pode representar também insatisfação do empregado com a empresa ou com o trabalho.

Avançando na definição do problema, Johns (2003) entendeu que o absenteísmo não é apenas um, mas uma grande variedade de comportamentos com diversas causas. Este fato se reflete, por exemplo, na multidisciplinaridade de conhecimentos disponíveis sobre o assunto, uma vez que esses conhecimentos tradicionalmente não se complementam, sendo desenvolvidos à parte por cada uma das disciplinas, que vão da psicologia à engenharia, passando por áreas como administração, enfermagem, ergonomia, direito e medicina, dentre outras.

Além desses fatores psicossociais causadores de absenteísmo, têm-se também como causas de absenteísmo as doenças e as dores, principalmente a dor de dente. Hito (2007) afirmou que os problemas odontológicos também se enquadram nos relatos da literatura científica que consideram o controle sobre o local de trabalho um importante predicativo para o absenteísmo. Porém, no atual contexto econômico sobre o absenteísmo, as empresas se preocupam em diminuir sua ocorrência e aumentar o rendimento na produção. Atuam no ponto final do processo, ou seja, no absenteísmo, e não nas causas que levam ao absenteísmo.

Segundo Hito (2007), o estudo do absenteísmo odontológico no Brasil é extremamente dificultoso por não haver uniformidade de documentos e não se guardarem arquivos referentes ao assunto, em detrimento dos estudos estatísticos.

Fiorelli (2008) declarou que a dor é um sinal de que algo no corpo não está bem; conflitos internos, falta de motivação, absenteísmo, rotatividade e queda na produtividade, são também sinais de alerta na empresa.

As características e o diagnóstico dos transtornos, orgânicos ou mentais, sofrem contínua transformação ao longo do tempo. Neste sentido, as causas de morte passaram de doenças causadas por infecção para doenças relacionadas ao comportamento e estilo de vida.

De acordo com Fiorelli (2008), o absenteísmo está relacionado com alterações psicológicas em função do processo de trabalho, notadamente em atividades que acarretam desgaste pessoal por excesso de responsabilidades.

Junkes (2010) afirmou que infecções virais, câncer e acidentes no ambiente de trabalho são parcialmente psicogênicos (fenômenos com origem psíquica), ou seja, são causados por condições emocionais ou psicológicas. O absenteísmo pode ser correlacionado com a raiva de colegas e o ambiente físico, esses fatores podem resultar em:

- redução da produtividade (aumenta a taxa de erros em procedimentos e diminui o ritmo de produção);
- conflitos interpessoais, entre pessoas da empresa e destas com os clientes;

- aumento do custo de vida (aquisição de medicamentos, consultas médicas, sensibilização para outros tipos de transtornos etc.), ocasionando maior necessidade de salários e benefícios, o que significa aumento com os custos diretos para a empresa.

Entretanto, o mascaramento de muitos transtornos mentais por sinais e sintomas somáticos faz com que se invistam grandes somas no combate a manifestações secundárias e não nas causas que levam aos transtornos mentais.

Os empregados afetados permanecem com qualidade de vida e produtividade reduzidas, em contínua busca de assistência médica e contribuição para aumentar os custos da medicina para toda a sociedade.

Em lugar de estimular a pessoa a buscar melhor qualidade de vida, a perturbação desencadeia reações (comportamentais, físicas e emocionais) percebidas como indesejáveis, crônicas ou agudas, acompanhadas de efeitos negativos no organismo, área física ou psíquica. Instala-se, então, o quadro conhecido como estresse (FIORELLI, 2008).

De acordo com Coleman (1992), as empresas perdem o conhecimento adquirido pelos empregados que desenvolvem quadro de transtorno de estresse, além de arcarem com os custos de reposição, tratamentos e perda de produtividade dos profissionais em conseqüências desses eventos. O empregado perde, em média, um ano e meio de trabalho durante a vida, devido a doenças provocadas pelo estresse.

2.1.4 Conseqüências do Absenteísmo nas Empresas

As conseqüências do absenteísmo estão diretamente ligadas ao lado financeiro das empresas, portanto, para toda e qualquer empresa, a produtividade e os lucros devem ser alcançados para que as empresas ocupem uma posição de destaque no mercado competitivo (AGUIAR e OLIVEIRA, 2009).

Segundo Fonseca (2009), a perda da produtividade associada com a saúde dos empregados de uma indústria automobilística de São Caetano do Sul, pode ser medida por meio do controle do absenteísmo, as conseqüências da ausência do empregado no trabalho são as faltas dispersas, as compensações e as ausências de curto e longo prazo geradas por incapacidades.

Quando se fala em saúde refere-se não apenas a saúde física, mas também os aspectos emocionais de uma pessoa. Os problemas emocionais, crises pessoais, alcoolismo, uso de drogas são problemas pessoais, que tornam problemas da empresa, pois, afetam o comportamento e o desempenho do empregado (MARTINS *et al.*, 2009).

Para Calgarotto e Pinheiro (2010), o absenteísmo consiste em fatores de desperdício e de perdas tanto para as empresas quanto para as pessoas, daí a importância de agir diretamente sobre as suas causas e não sobre seus efeitos, sendo assim, necessário diagnosticar o Absenteísmo.

2.1.5 O Custo do Absenteísmo

O custo do absenteísmo por motivo de doença é, em muitas empresas, maior do que o salário pago ao empregado doente que deixa de comparecer ao local de trabalho e, não havendo substitutos disponíveis, as implicações para a produtividade são bem maiores. Um serviço deixa de ser fornecido devido à falta do empregado, reduzindo a receita da empresa (CHIAVENATO, 2004; JUNKES, 2010; PAULY, 2012).

A perda da produtividade do empregado resultante de problemas de saúde é uma medida indireta do custo de saúde para as empresas e é geralmente avaliada pelos índices de absenteísmo e medidas de desempenho do trabalho, sendo geralmente descritas como de difícil quantificação (FONSECA, 2009).

Danatro (1997) ressaltou num estudo feito numa instituição pública de Montevideo, que o absenteísmo representa mais de 10% de todas as despesas de todos os salários e de todas as horas extras pagos ao empregado.

Muitas empresas perceberam que as ausências por doença resultam em perda da força de trabalho, pagamento para empregados temporários, redução da produtividade e aumento da rotatividade de empregados (SANTOS, 2008).

O absenteísmo gera um custo anual estimado em mais de 40 bilhões de dólares nas empresas norte-americanas. Além do prejuízo financeiro, o absenteísmo é um fator que pode comprometer fatalmente o sucesso de uma empresa, podendo levá-la a falência (TAUFFER e COLTRE, 2007). Para Mallada (2010), o absenteísmo sempre causa custo para empresa, porque o empregado, em tais circunstâncias, continua a receber sua remuneração.

De acordo com Cunha *et al.* (2009), a Fundação Europeia para a Melhoria das Condições de Vida e Trabalho destacou os elevados custos diretos e indiretos do absenteísmo decorrente de faltas por doença, atribuídos a perdas na produtividade e, também, os custos diretos dos afastamentos impactam severamente nas contas de outros países.

Junkes (2010) afirma que são considerados custos ocultos a perda de motivação, caso em que normalmente o empregado é vítima de estresse nos relacionamentos com pessoas difíceis, principalmente sob assédio moral por suas chefias.

Em Rocha (2010), as consequências psicossociais no absenteísmo por doença, os empregados arcam com 30% do total da despesa resultante do absenteísmo.

De acordo com Cunha *et al.* (2009), a pesquisa com o 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.

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 *et al.*, 2006; CUNHA *et al.*, 2009).

De acordo com Rocha (2010), devido aos altos custos do absenteísmo, as empresas procuram administrá-lo de modo que tenha que reduzir ao máximo suas perdas. Este tem sido um dos maiores desafios para o setor de saúde ocupacional, dada a dimensão do fenômeno que engloba a detecção e monitoramento das causas de fundo psicológico, físicos e ambientais, considerando as circunstâncias que levam o empregado ao absenteísmo.

2.1.6 Quantificação do Índice de Absenteísmo

Quando o absenteísmo é medido em função do tempo total perdido (somatório de dias ou horas, ausências prolongadas), o mesmo é considerado involuntário. Por outro lado, a medida da frequência do comportamento (ausências repetidas) evidencia o absenteísmo voluntário (DIAS, 2005).

Assim, segundo Marras (2000), o absenteísmo é controlado através do índice de absenteísmo, que tem como finalidade subsidiar análises de variações e tendências de horas perdidas de trabalho. O índice de absenteísmo, possibilita que a empresa perceba então, que consequências as ausências dos seus empregados trazem para os resultados da mesma, bem como para os próprios empregados.

Existem diversas formas para se quantificar o índice de absenteísmo, cada empresa utiliza a forma que mais se adequa à sua realidade e a finalidade que se pretende utilizar. A seguir, é apresentada a Equação (2.1) (CHIAVENATO, 1997; MARRAS, 2000):

$$Ia = \frac{Nh_p}{NhP} \cdot 100 \quad (2.1)$$

Onde:

Ia = índice de Absenteísmo

Nh_p = Número de horas perdidas

NhP = Número de horas planejadas

Chiavenato (1997), traz duas possibilidades de cálculo para a quantificação do índice de absenteísmo, conforme apresentado nas Equações (2.2) e (2.3). Na primeira, são considerados os dias de trabalho perdidos:

$$Ia = \frac{N_e}{d_p} \cdot 100 \quad (2.2)$$

Onde:

Ia = índice de Absenteísmo

N_e = número de empregados

d_p = dias perdidos por ausência ao trabalho

A segunda possibilidade de quantificação do índice de absenteísmo exposta pelo autor, não considera os dias de trabalho perdidos, mas sim as horas, já que, segundo este autor, o absenteísmo também pode ocorrer por meio de atrasos, onde o sujeito se ausenta por algumas horas ou por algum período do dia.

$$Ia = \frac{T_e \div h_p}{T_e \div h_t} \cdot 100 \quad (2.3)$$

Onde:

Ia = índice de Absenteísmo

T_e = total de empregados

h_p = horas perdidas

h_t = horas trabalhadas

As formas de cálculo de absenteísmo que consideram as horas perdidas podem ser consideradas mais apropriadas do que as formas de cálculo por dia, já que os atrasos são tão significativos quanto às faltas do empregado à sua função. Para Gehring Junior *et al.* (2007), existem diversos motivos que podem levar o empregado a abster-se do trabalho, tais como problemas de saúde ou doenças, acidentes do trabalho, problemas de saúde em pessoas da

família, licença maternidade e paternidade, casamento (licença gala), entre outros. Podem ocorrer ainda faltas que não são justificadas legalmente, conhecidas como faltas injustificadas.

A preocupação com a QVT é algo que está ligada à produtividade da empresa. Empregados satisfeitos e motivados são mais comprometidos com o trabalho e condições ótimas de saúde aliada a um ambiente saudável leva à qualidade produtiva pretendida (MIRANDA *et al.*, 2006).

Bonventti (2012) destacou que as empresas ganharam em termo de competitividade no mercado quando estimularam uma cultura preventiva, promoveram a melhoria da saúde tanto do empregado quanto de seus dependentes e reconheceram que, o melhor para a redução do absenteísmo prolongado por motivo de doença, é justamente o uso dos benefícios da assistência médica como medidas preventivas das doenças que afetam seus empregados, além de proporcionar qualidade de vida no trabalho e reduzir significativamente os custos com afastamentos médicos.

Em paralelo, políticas de recursos humanos e dos gestores devem ser cada vez mais ativos no controle do absenteísmo, pois os resultados individuais parecem ser influenciados significativamente através de incentivos, reconhecimento, preocupação com o bem-estar dos empregados e o atendimento a suas necessidades para o desempenho adequado (DIAS, 2005). Isso permite reconciliar os objetivos dos indivíduos em situação de trabalho e os propostos pelas empresas.

2.2 O PRESENTEÍSMO

Em paralelo ao absenteísmo, foi identificado, nos últimos anos, um novo conceito: o presenteísmo.

Para 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 conceito de presenteísmo iniciou na década de 50, quando os pesquisadores Canfield e Soash (1955) e Uris (1955) começaram a buscar as causas do absenteísmo e doenças da natureza do trabalho. O presenteísmo tornou-se mais conhecido durante os últimos anos do século XX, quando as empresas perceberam que a barreira da competitividade está nas mãos dos empregados (GOETZEL *et al.*, 2004; O' DONNELL, 2009).

De acordo com 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 e redução dos benefícios.

Para Martins *et al.* (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.

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

Neste aspecto, para as empresas, que querem permanecer no mercado competitivo, é de suma importância um planejamento voltado para o combate dos fatores do absentéismo e do presenteísmo, haja vista a relação evidente entre esses fatores e a baixa qualidade dos serviços prestados nos diferentes ramos de produção (DALL'INHA, 2006).

Hemp (2004) e O' Donnell (2009) definiram o presenteísmo como um conceito para explicar o fato dos empregados estarem no local de trabalho, mas, devido a problemas de ordem física ou psicológica, não conseguirem cumprir totalmente as suas funções.

Levin-Epstein (2005) definiu o presenteísmo como perda de produtividade, que ocorrem quando os empregados estão no trabalho, porém com rendimento reduzido devido algum tipo de doença. Enquanto o absentéismo tem sido muito pesquisado, o mesmo não ocorre com o Presenteísmo.

Martinez *et al.* (2007) concordam 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) psico-somático: 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, distúrbios gastrointestinais, artrites e depressão (PASCHOALINO, 2008).

Segundo o mesmo autor, presenteísmo significa estar sempre presente no local de trabalho, porém doente. Esses profissionais que estão apresentando sintomas do presenteísmo em suas condutas colocam o trabalho em primeiro lugar, esquecendo-se, inclusive, de sua saúde. A primeira consequência do presenteísmo está vinculada à perda da produtividade frente ao esforço do profissional em manter-se doente no trabalho.

De acordo com Laranjeira (2009), o conceito foi usado para designar o fenômeno em que as pessoas vão trabalhar, mas realizam as atividades inerentes às suas funções de um modo não produtivo, ou seja, não apresentam um bom desempenho, porque apresentam problemas de saúde físicos e mentais.

2.2.1 O Presenteísmo nas Empresas

Hoje, a competição entre os empregados é algo que ocorre com muita frequência dentro das empresas, estes empregados podem ser considerados como *Horkaholic*, pessoa cuja vida, nos casos mais graves, pode reduzir-se ao trabalho. Ele pode ser considerado um presenteísta, já que nem sempre sua produtividade e eficiência correspondem ao elevado número de horas trabalhadas (GOETZEL, 2004).

É importante observar que dentro das empresas existem empregados que vão trabalhar mesmo quando estão doentes, o que caracteriza o presenteísmo que, de acordo com Primo (2008), pode estar presentes nos locais onde o índice de absentéismo é baixo.

O presenteísmo também é determinado pela cultura organizacional, em especial no que diz respeito ao papel simbólico da presença no local de trabalho. Nas empresas, onde a cultura da presença simbólica tem um valor elevado, os empregados disputam entre si para ver quem fica mais tempo no ambiente de trabalho. Gera-se desta forma um presenteísmo competitivo entre os empregados, o que acarreta vários problemas de saúde que resultam em absentéismo (MARTINEZ *et al.*, 2007). Isso afeta negativamente a produtividade dos empregados, além de ser considerado um dos sintomas que antecede o estresse ou a depressão.

Segundo Rocha (2010) o empregado que trabalha doente tem maior chance de se afastar por longo prazo, podendo apresentar mudanças na qualidade e expectativa de vida.

2.2.2 As Causas do Presenteísmo

As causas de presenteísmo devem-se a problemas psicológicos - depressão, estresse e exaustão emocional - músculo-esquelético - lombalgias, artrites - e respiratório - síndromes gripais, asma entre outras causas (LARANJEIRA, 2009).

O estresse é um tema bastante difundido nos dias de hoje, pois leva, em sua essência, problemas de adaptação a situações de ansiedade. O excesso de fatores estressores presentes no ambiente de trabalho faz com que haja a necessidade de mobilizar mais energia do que realmente se dispõe, que ultrapassam a capacidade física ou psíquica do indivíduo. Níveis extremos de estresse ocorrem devido a um acontecimento emocional negativo gerado por sensação de medo, tensão, derrota, raiva, cansaço e falta de iniciativa, isso pode ocasionar a falta de adaptação do homem ao seu ambiente de trabalho (COSTA e COSTA, 2003).

O estresse é definido por Pafaro e Martino (2004) como um desgaste geral do organismo, causado pelas alterações psicofisiológicas, que ocorrem quando o indivíduo é forçado a enfrentar situações que o irrite, excitem, amedrontem.

Transtornos pouco, ou menos, presentes há algumas décadas acentuam-se, enquanto outros apresentam redução. Por exemplo, registra-se perceptível aumento de transtornos associados a estresse (FIORELLI, 2008).

Para Primo (2008), o estresse se converteu numa das principais causas de afastamento em vários seguimentos laborais. A manifestação desse fenômeno pode ser atribuída às exigências da organização do trabalho. As repercussões desse fenômeno do ponto de vista da perda de produtividade, absenteísmo, perda da qualidade de vida e adoecimento são enormes e de difícil mensuração, por tratar-se de risco, muitas vezes, não reconhecido ou declarado.

Segundo Fiorelli (2008), além das doenças relatadas em decorrência do estresse, pode-se observar outros fatores de diferentes origens e que levam ao presenteísmo:

- frustração, pelo fracasso na busca de um objetivo;
- conflito, decorrente da competição entre motivações ou impulsos comportamentais incompatíveis;
- mudança, resultante de uma alteração na forma de vida da pessoa;
- pressão, envolvendo expectativas ou exigências para que a pessoa se comporte de determinada forma.

O presenteísmo limita a produtividade não só em termos de quantidade de trabalho, mas também em termos de qualidade do trabalho produzido (HEMP, 2004; MARTINEZ *et al.*, 2007).

As repercussões da qualidade dos serviços prestados refletem-se em erros e omissões nos procedimentos de trabalho enquanto que, no que diz respeito à quantidade, a produção não corresponde aos objetivos esperados, devido a dificuldades de concentração. A queda da produtividade traduz-se assim na incapacidade de desempenhar as tarefas de rotina de trabalho, o que pode resultar em elevadas taxas de abandono por doença, que justificam alguma relação entre Presenteísmo e Absenteísmo (MARTINEZ *et al.*, 2007).

Laranjeira (2009) definiu o estresse como a interação do empregado com as condições de trabalho, de tal modo, que as exigências resultantes desta interação ultrapassam a capacidade do empregado para lidar com a situação. Isso não tem somente características cumulativas, como torna o ser humano mais sensível aos acontecimentos adversos com que se depara, o que por vezes se traduz em implicações negativas no ambiente familiar e do trabalho.

Para Umann (2011), o estresse pode ser definido como uma reação inespecífica do organismo, síndrome da adaptação geral ou do estresse biológico caracterizada como uma reação defensiva fisiológica do organismo em resposta a qualquer estímulo aversivo.

Além dos vários sintomas que podem levar o empregado ao presenteísmo, citados na seção 2.2, o sintoma com maior impacto e considerado fator de risco à saúde dos empregados é o estresse, que é o maior gerador do Presenteísmo. A Figura 2.1 mostra os países mais estressados do mundo em ordem decrescente:

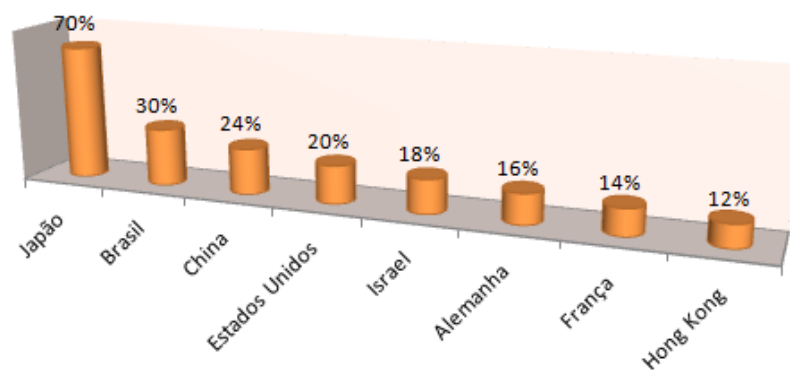


Figura 2.1 – Países mais estressados do mundo. Fonte: Fava (2011)

Com base na Figura 2.1 percebe-se que o estresse no Brasil, está classificado como o segundo país mais estressado do mundo perdendo apenas para o Japão que apresenta o maior nível de estresse com 70%. A China aparece em terceiro lugar com 24% e os Estados Unidos em quarto lugar com 20%, seguido por Israel com 18%, Alemanha com 16%, França com 14% e Hong Kong com menor nível de estresse entre os oito países pesquisados, apenas 12% (FAVA, 2011).

No Brasil, 30% da população economicamente ativa sofre com o estresse no ambiente de trabalho. As principais causas são a sobrecarga de tarefas e o medo da demissão.

Na mesma direção, percebe-se que, devido ao processo de inserção cada vez maior de empresas nacionais no cenário competitivo mundial, os profissionais desse meio já sentem os efeitos de fatores estressores, outrora exclusivas dos países industrializados (COSTA e COSTA, 2003).

2.2.3 O Custo do Presenteísmo

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).

Ao analisarem os custos diretos e indiretos das despesas com saúde, Goetzel *et al.* (2002), Hemp (2004), Martinez *et al.* (2007) e O' Donnel (2009) estimaram perdas de produtividade anuais na ordem dos 180 mil milhões de dólares, incluindo absenteísmo e presenteísmo. O estado fisiológico e psicológico dos empregados é, por isso, não só importante em termos de responsabilidade social da empresa, mas também em termos da sua competitividade (MARTINEZ *et al.*, 2007).

Estima-se que as empresas brasileiras gastem anualmente 12,5 bilhões de reais com despesas decorrentes de acidentes e doenças relacionadas ao trabalho e que os contribuintes gastem mais de 20 bilhões de reais anuais com estes tipos de despesas. Tais custos salientam a necessidade de programas de prevenção às doenças ocupacionais, dentre os quais o estresse ocupacional, considerado fator de risco para várias doenças ligadas ao trabalho (MURTA *et al.*, 2005).

A distribuição dos custos do presenteísmo de acordo com as condições médicas também tem sido foco de estudos que estimam os custos totais do presenteísmo. A Figura 2.2 mostra a estimativa dos custos totais de presenteísmo para as condições específicas de saúde.

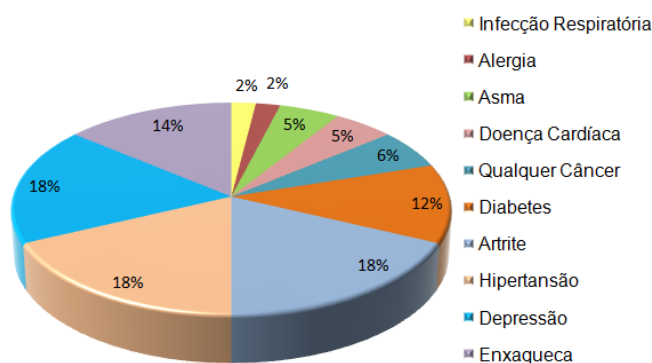


Figura 2.2 – Estimativa sobre os custos do presenteísmo e as condições específicas de saúde.

Fonte: Adaptado de Goetzal *et. al* (2004)

Conforme a Figura 2.2, as cinco condições de saúde com os custos mais elevados são: a depressão, hipertensão, artrite, enxaqueca e diabetes. Estas condições são responsáveis por aproximadamente 75% dos custos do presenteísmo.

2.2.4 A Influência do Presenteísmo no Absenteísmo

As empresas que direcionam medidas excessivas à redução do absenteísmo, ocasionam um aumento de presenteísmo, que a longo prazo podem aumentar o afastamento por doenças (absenteísmo prolongado) (BERGSTRON *et al.*, 2009).

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.

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.

Na Figura 2.3, as despesas médicas e farmacêuticas correspondem aos pagamentos feitos aos empregados, pedidos de tratamento médicos e farmacêuticos prescritos. As despesas de incapacidade e absenteísmo foram representadas pelo valor pago quando os empregados estão longe do trabalho (HEMP, 2004).

As despesas com o presenteísmo foram estimadas nos salários dos empregados e correspondem aos valores perdidos para a redução de doenças relacionadas com a produtividade (Figura 2.3).

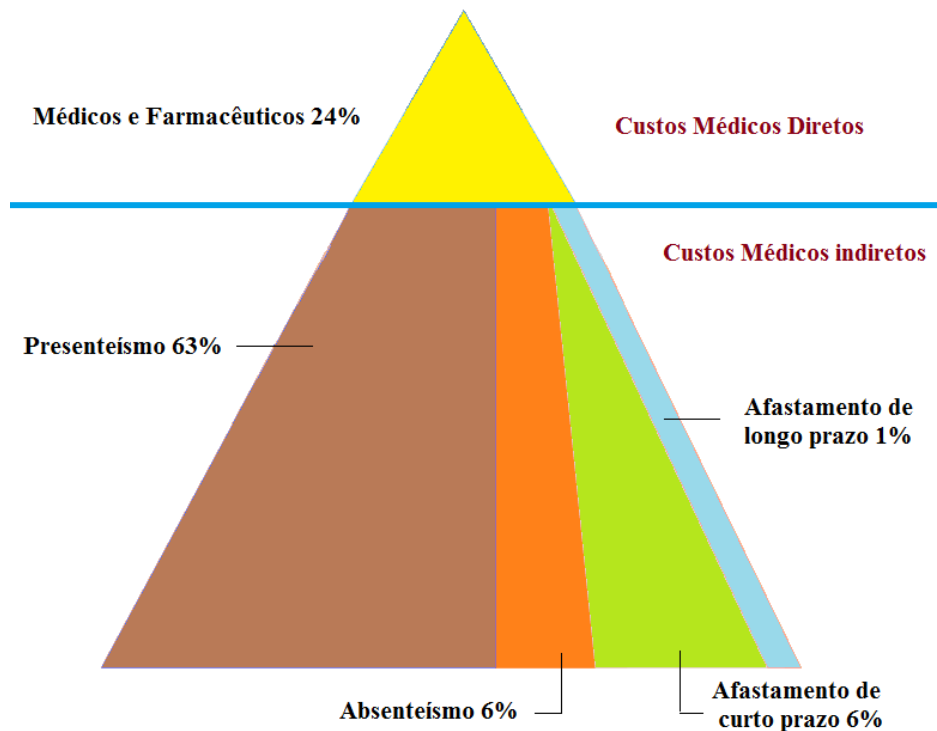


Figura 2.3 – Custos com a saúde ocupacional. Fonte: Adaptado de Hemp (2004)

Observa-se na Figura 2.3 que 63% dos custos médicos indiretos com a saúde estão relacionados com o presenteísmo e 13% estão relacionados com o absenteísmo.

O que pode parecer controverso para empresa, é que o presenteísmo pode custar substancialmente mais do que os custos diretos nos tratamentos médicos e farmacêuticos, 24% (HEMP, 2004).

Estes dados refletem 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, gerando prejuízos econômicos significativos para empresa e compromete a estabilidade da empresa no mercado de trabalho competitivo (MCKEVITT *et al.*, 1997).

2.2.5 Quantificação do Índice de Presenteísmo

De acordo com Hemp (2004), as empresas pesquisam para determinar a prevalência de doenças e outras condições médicas que prejudicam o desempenho do trabalho; procuram calcular a queda de produtividade e encontrar formas rentáveis de luta contra essa perda; e descobrem que os declínios do presenteísmo, relacionadas com a produtividade, podem ser compensados com investimentos relativamente pequenos na triagem, no tratamento e na educação.

A quantificação do presenteísmo segundo Mettike *et al.* (2007) ainda está em processo de desenvolvimento. Quanto aos instrumentos usados para medir o presenteísmo, faltam métodos estabelecidos e validados para converter a perda de produtividade em estimativas monetárias. Os desafios para medir o presenteísmo excedem de longe os de medir o absenteísmo, sobretudo porque em muitas profissões os seus resultados não são facilmente mensuráveis (MATTIKE *et al.*, 2007).

Antes de quantificar o presenteísmo as empresas devem estabelecer o controle da carga de trabalho, bem como as fases de desenvolvimento desse trabalho, sendo que estas são partes cruciais do planejamento da produção e do sistema de controle (WU e WENG, 2005).

De acordo com Martinez *et al.* (2007), há um instrumento que permite a avaliação do Presenteísmo, tendo a escolha recaído no *Stanford Presenteeism Scale-6* (SPS-6). A versão original deste instrumento foi desenvolvida e validada por Koopman e colaboradores em 2002.

A escala do SPS-6 é constituída por 6 afirmações baseadas numa escala do tipo-Likert com cinco níveis de resposta (de 1 – discordo totalmente, a 5 – concordo totalmente). Esta é uma das mais interessantes medidas disponíveis para estimar as perdas de produtividade laboral e tem sido um dos instrumentos mais utilizados para medir o presenteísmo. A estrutura da escala integra duas dimensões de presenteísmo, com três itens cada. A primeira encontra-se associada à quantidade de trabalho, que é efetuado quando o empregado está sob a influência das causas do presentismo. A segunda diz respeito à capacidade de concentração das pessoas quando se manifestam os sintomas do presenteísmo (MARTINEZ *et al.*, 2007).

Segundo Thorwarth *et al.* (2009), um número crescente de pacotes de *softwares* de simulação podem ser aplicados ao processo de otimização. Entretanto, suas aplicações concentram-se na fabricação e instalação de produção e, raramente, tratam da dinâmica do

trabalho, tais como carga de trabalho e flexibilidade das tarefas dos empregados, em que verifica-se menor atenção.

Este modelo pode ser utilizado não apenas para medir a qualidade dos serviços prestados, mas pode ser um direcionador dos custos para identificar o número adequado de empregados dentro de uma empresa para cada atividade e horários diferentes de trabalho para se manter um equilíbrio, em que o estresse, causador do presenteísmo, é reduzido, bem como são reduzidas as consequências negativas do absenteísmo e conseqüentemente os custos da produção (THORWARTH *et al.*, 2009).

2.3 DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

A Descoberta de Conhecimento em base de Dados ou *Knowledge Discovery in Databases* (KDD) foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de bases de dados. De acordo com Fayyad *et al.* (1996a), 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.

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 e 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 e 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, a mineração de dados ou *Data Mining* (DM), 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. A mineração de dados, 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 e CAMPOS, 2006).

2.3.1 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 *et al.*, 1996a). 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 e
- e) Interpretação/avaliação do conhecimento.

Para ilustrar as fases do processo KDD segue a Figura 2.4, representando suas várias etapas.

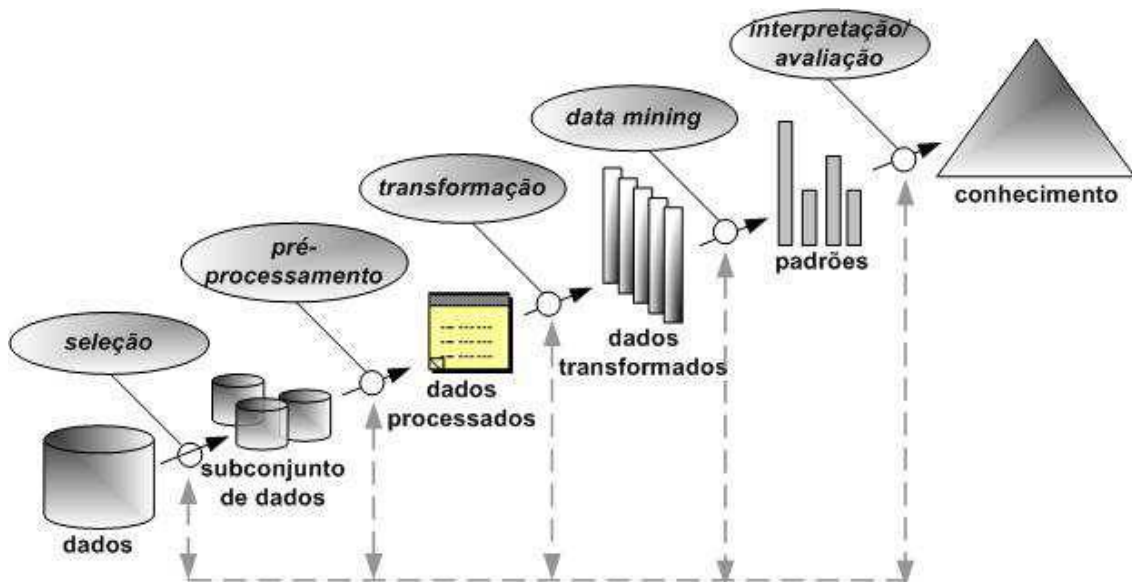


Figura 2.4 – Fases do processo de KDD. Fonte: Adaptado de Fayyad *et al.* (1996a) e de Sassi (2006)

O processo compreende todo o ciclo que o dado percorre até gerar conhecimento. A iteração entre as fases pode ser observada pelas setas tracejadas (Figura 2.4).

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 e Campos (2006), as fases do KDD, que antecedem a mineração de dados, 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.

A Figura 2.5 ilustra o tempo necessário para cada fase do processo KDD.

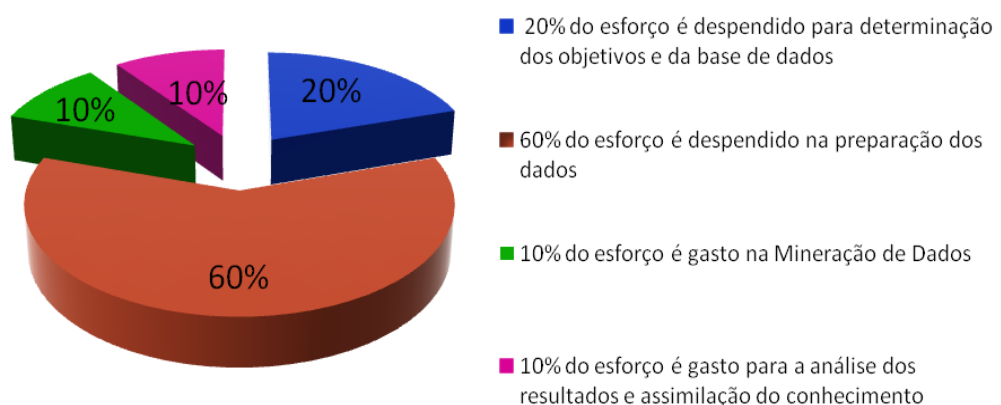


Figura 2.5 - Esforço requerido para cada fase do processo do KDD. Fonte: Adaptado de Cabena *et al.* (1998) e Sassi (2006)

Observa-se na Figura 2.5, que 60% do tempo necessário é requerido na preparação dos dados (seleção, pré-processamento e transformação), visando organizar os dados para facilitar o trabalho realizado pelas fases posteriores do processo KDD com maior qualidade. As fases do KDD são explicitadas a seguir:

a) Seleção de Dados

A fase de seleção de dados é a primeira no processo de descobrimento de informação. Nesta fase, busca-se determinar o subconjunto de atributos sobre o qual o KDD deverá atuar, facilitando o trabalho dos algoritmos responsáveis pela tarefa de organização e ordenação dos dados. Escolhem-se apenas atributos relevantes do conjunto de atributos da base de dados, que estejam de acordo com os objetivos da empresa. O subconjunto selecionado é então fornecido para o algoritmo de mineração dos dados até a transformação dos dados para torná-lo mais compreensíveis.

O processo de seleção é bastante completo, uma vez que os dados podem vir de uma série de fontes e formatos. O tempo de processamento do algoritmo minerador, visto que ele apenas trabalhará com um subconjunto de atributos, é uma motivação para essa seleção, diminuindo o seu espaço de busca (BARROS e CAMPOS, 2006; SASSI, 2006).

b) Pré-processamento e Limpeza dos Dados

Nesta etapa de seleção, limpeza e tratamento dos dados são realizadas tarefas que excluem dados redundantes e inconsistentes, recuperam dados incompletos e avaliam possíveis dados divergentes ao conjunto (*outliers*), o que permite a qualidade dos dados selecionados, o auxílio do especialista do domínio é fundamental, pois é o mesmo que definirá se os padrões adquiridos são interessantes, se o conhecimento é válido, novo e útil, ou se será necessário retornar a alguma das etapas anteriores (ANUMALLA, 2007).

Nesta fase também são utilizados métodos de redução de dados ou transformação para diminuir o número de variáveis envolvidas no processo, compreende a identificação de quais informações, dentre as bases de dados existentes, devem ser consideradas durante o processo de KDD, visando com isto melhorar o desempenho do algoritmo de análise (FAYYAD *et al.*, 1996b; GOLDSCHMIDT e PASSOS, 2005).

A limpeza dos dados envolve a verificação de ruídos, dados estranhos ou inconsistentes, e são estabelecidas as estratégias para resolver os problemas de ausências de dados. A execução dessa fase corrige a base de dados eliminando consultas desnecessárias,

que seriam executadas pelo algoritmo minerador e que afetariam o seu processamento (SASSI, 2006; SILVA NETO, 2008).

c) Transformação de Dados

A fase de transformação de dados ou codificação dos dados tem como objetivo principal converter o conjunto bruto de dados em uma forma padrão de uso. (GOLDSCHMIDT e PASSOS, 2005).

Os dados de um atributo podem ser padronizados (normalizados) para ficar dentro de uma faixa de valores, como por exemplo: -1,0 a 1,0 ou 0,0 a 1,0.

A transformação de dados para o atributo idade poderia ocorrer da seguinte forma: {0...15}. Faixa 1; {16...20}. Faixa 2; {21...30}. Faixa 3 e assim por diante (DOUGHERTY *et al.*, 1995). Nesse exemplo os valores contínuos das idades foram discretizados em faixas. Em resumo, essa fase converte os dados para atender as necessidades mais adequadas para a construção do modelo.

A transformação de dados é potencialmente a tarefa que requer grande habilidade no processo de KDD. Normalmente, essa etapa exige a experiência do analista de dados e seu conhecimento nos dados em questão.

Embora o processo de KDD possa ser executado sem essa fase, nota-se que quando efetivada os resultados obtidos são mais intuitivos e valiosos, além de, na maioria das vezes, facilitarem a construção do modelo.

As vantagens de se codificar um atributo são melhorar a compreensão do conhecimento descoberto, reduzir o tempo de processamento para o algoritmo minerador e facilitar o algoritmo a tomar decisões globais, já que os valores dos atributos foram englobados em faixas. Como desvantagem cita-se a redução da medida de qualidade de um conhecimento descoberto, perdendo-se assim detalhes relevantes sobre as informações extraídas.

d) Mineração de Dados

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

Segundo Carvalho (2005), as ferramentas da nova ciência dita mineração de dados 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 mineração de dados 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.

Durante a etapa de mineração de dados é realizada a busca efetiva por conhecimento útil no contexto da aplicação do KDD. Onde são definidas as técnicas e os algoritmos a serem utilizados no problema em questão. A escolha da técnica depende, muitas vezes, do tipo de tarefa do KDD a ser realizada (GOLDSCHMIDT e PASSOS, 2005).

De acordo com Barros e Campos (2006), mineração de dados trata-se de um conjunto de técnicas reunidas da Estatística e da Inteligência Artificial com o objetivo específico de descobrir conhecimento novo, que porventura esteja escondido em grandes massas de dados armazenadas em bancos de dados.

Rodrigues (2001) já explicava que a mineração de dados procura uma série de padrões escondidos nos dados, envolve uma aplicação iterativa e repetitiva de métodos de mineração de dados. Com o auxílio da mineração de dados, podem-se desenvolver aplicações que mostrem de maneira gráfica informações críticas, com o objetivo de estreitar ainda mais o relacionamento das empresas com seus consumidores finais ou a identificação, também em modo gráfico, de problemas da produção que acabam passando despercebidos.

Para Costa (2007), os dados produzidos e armazenados em larga escala possuem alta complexidade em sua leitura e análise por especialistas humanos. A partir dessa visão fez-se necessário criar uma nova geração de métodos e técnicas capazes de auxiliar o ser humano a buscar conhecimento útil nas bases de dados. A mineração de dados surgiu como uma das principais soluções para auxiliar o processo de KDD.

Segundo Piatetsky-Shapiro *et al.* (1993), a mineração de dados serve para prever tendências e comportamentos futuros, o que permite a tomada de decisão baseada em fatos e não em suposições, que exigiriam muito tempo para resolver. Além disso, ela usa técnicas baseadas em descobertas por meio de procura de padrões dos dados, utilizando uma série de algoritmos inteligentes para encontrar relações fundamentais entre os dados.

Essas técnicas e algoritmos caracterizam-se por serem "automáticos", ou seja, são de baixa influência do ser humano, a participação humana se restringe aos processos de ajuste necessários aos mecanismos de busca (PIATETSKY-SHAPIRO *et al.*, 1993). Para compreender a busca pelo conhecimento, ele passa por algumas etapas até atingir a interpretação. A Figura 2.6 mostra um esquema simplificado da mineração de dados.

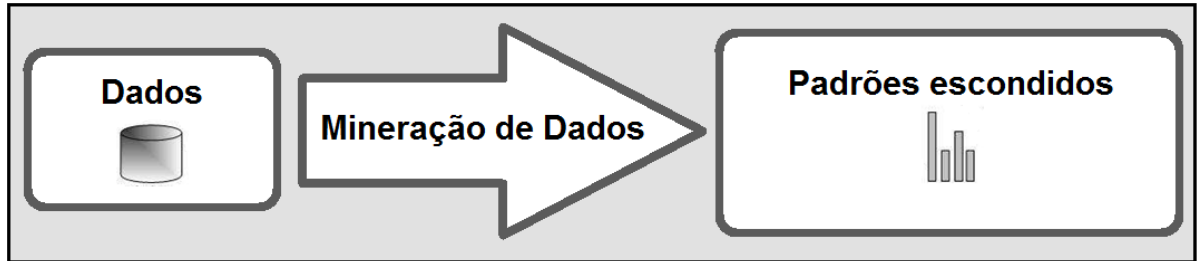


Figura 2.6 – Esquema simplificado da mineração de dados. Fonte: Adaptado de Sassi (2006)

Esse processo inicia-se com os dados que são os primeiros reflexos do saber, que são transformados em informações válidas, ou seja, o conjunto de dados é codificado. Após essas etapas, chega-se a interpretação dos dados coletados e transformados em conhecimento através da mineração de dados para auxiliar nas decisões e no crescimento organizacional (COSTA, 2007).

A qualidade do conhecimento extraído deve ser considerada uma questão indispensável da mineração de dados e deve-se levar em conta a precisão, a compreensibilidade, a surpresa e/ou o potencial interesse do conhecimento obtido (SASSI, 2006).

e) Interpretação/Avaliação do conhecimento

A fase de transformação é implementada por meio de um processamento dos dados, visando organizar os dados para facilitar o trabalho realizado pelas fases posteriores do processo de KDD (BARROS e CAMPOS, 2006). A interpretação da informação extraída é analisada em relação ao objetivo do auxílio à decisão do usuário final, sendo os melhores resultados das informações identificadas e expostas. Caso o resultado não seja satisfatório, o processo pode retornar a qualquer um dos estágios anteriores.

A Interpretação/Avaliação do conhecimento também é conhecida como pós-processamento, pois atua no conjunto de regras descoberto pela fase de mineração de dados com o intuito de garantir um bom grau de compreensão do conhecimento descoberto pelo algoritmo minerador, validando-o através de medidas da qualidade da solução e da percepção do analista de dados em medir o quão novo e surpreendente é o conhecimento descoberto. Uma forma genérica de atingir o entendimento e avaliação dos resultados é utilizar técnicas de visualização (BIGUS, 1996).

As técnicas de visualização incentivam a percepção e a inteligência humana, ampliam a capacidade de compreensão e a associação de novos padrões (GOLDSCHMIDT e PASSOS, 2005).

2.3.2 Tarefas de KDD

Existem várias formas de interpretação dos dados pela KDD denominadas tarefas. Diversas tarefas podem ser realizadas, as tarefas mais comuns segundo Fayyad *et al.* (1996a) são associação, classificação, clusterização (agrupamento) e visualização.

a) Associação

De acordo com Goldschmidt e 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, 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 e 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 e 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 e 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, 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 *et al.* (1996a), 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 *et al.*, 1996a; GOLDSCHMIDT e PASSOS, 2005).

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 mineração de dados 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 mineração de dados. No caso específico da rede SOM, as formas de visualização mais utilizadas são a Matriz-U (ULTSCH e SIEMON, 1989) e o Mapa por Similaridade de Cor.

As ferramentas de visualização podem ser usadas para auxiliar, ou serem auxiliadas, pelas técnicas de mineração de dados. 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).

2.4 O MAPA AUTO-ORGANIZÁVEL DE KOHONEN

Em contraste às máquinas computacionais tradicionais que se baseiam em processadores relativamente complexos, as Redes Neurais Artificiais (RNAs) consistem em um grande número de unidades computacionais simples, chamadas de neurônios altamente interconectadas com diversas vias de comunicação (SIMÕES e SHAW, 2007).

A Figura 2.7 ilustra a representação do neurônio biológico e do neurônio artificial.

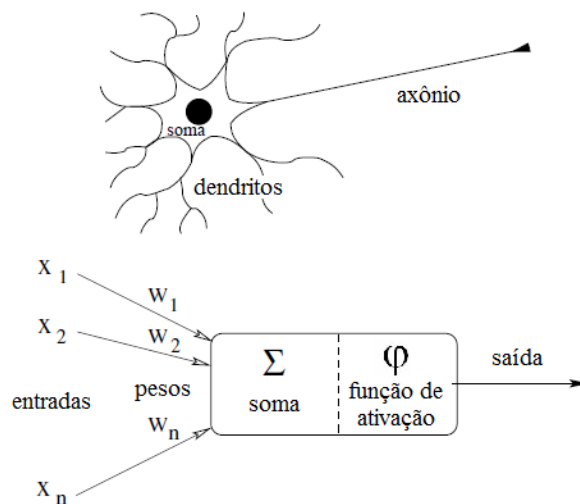


Figura 2.7- Representação do neurônio biológico e artificial. Fonte: Adaptado de Paugam-Moisy e Bohte (2010)

Pode-se perceber a semelhança entre as duas estruturas, as entradas de um neurônio artificial representam os dendritos do neurônio biológico, a junção aditiva no neurônio artificial representa o corpo celular no neurônio biológico e a saída do neurônio artificial representa o axônio no neurônio biológico (BRAGA *et al.*, 2011).

O neurônio artificial, similarmente ao natural, recebe um ou mais sinais de entrada e devolve um único sinal de saída, que pode ser distribuído como sinal de saída da rede, ou como sinal de entrada para um ou vários outros neurônios da camada posterior. Os sinais de entrada chegam simultaneamente aos neurônios, por isso as redes são classificadas como instrumentos de processamento paralelo, que em máquinas sequenciais, é simulado pela execução de uma varredura através da rede (HAYKIN, 2001).

Em suma, uma RNA é uma tradução simplificada do funcionamento do cérebro humano e do sistema nervoso, a partir de algoritmos. As RNAs caracterizam-se por possuírem elevado número de elementos de processamento simples inspirados no funcionamento do neurônio biológico, conexões entre os elementos de processamento, um peso associado em cada conexão, que representa quão forte é a interação, ou acoplamento, entre os elementos de processamento e se a sua natureza é excitatória ou inibitória (MAASS e BISHOP, 2001; BRAGA *et al.*, 2011).

O início das pesquisas sobre a neurocomputação ou modelagem neural se deu com a proposta em um artigo de McCulloch e Pitts (1943), no qual propuseram um modelo de um neurônio artificial inspirado no neurônio biológico.

Em Hebb (1949), é proposto o primeiro método de treinamento para RNAs, denominado regra de aprendizado de Hebb, baseada em hipóteses e observações de caráter neurofisiológico.

No final da década de 50, Frank Rosenblatt deu prosseguimento às idéias de McCulloch e Pitts e criou uma Rede Neural Artificial, denominada *Perceptron* de múltiplas camadas ou *Multilayer Perceptron* (MLP). Esta rede, cuja origem era uma simulação computacional para a retina, demonstrou como o sistema nervoso visual reconhece padrões simples, esse modelo despertou interesse e, em 1960, Windrow e Hoff utilizaram o conceito para formular o *Adaline* (*Adaptive Linear Element*), uma rede cujo aprendizado é baseada na regra Delta (KOVÁCS, 2006).

Em 1969, Marvin Minsky e Seymour Papert publicaram o livro intitulado *Perceptron*, provando que as RNAs de uma única camada de Rosenblatt eram incapazes de resolver alguns problemas simples, como por exemplo, o do Ou-exclusivo e demonstraram as limitações de *Perceptrons* de uma camada.

Esse fato, aliado a dificuldades de implementação no *hardware* ou *software* então disponível, levou muitos pesquisadores a diminuírem as suas atividades nesse campo (NASCIMENTO JÚNIOR e YONEYAMA, 2004; PACHECO e VELLASCO, 2007).

Na década de 1980, foram feitas importantes contribuições em várias frentes à teoria e ao projeto de Redes Neurais Artificiais e, com isso, houve um ressurgimento do interesse pelas RNAs.

Foram desenvolvidos trabalhos nas áreas de memória associativa endereçável pelo conteúdo, sistema visual e reformulação do problema de aprendizado das redes. Em Grossberg (1980), foram desenvolvidos novos princípios de auto-organização e aprendizado competitivo, originando uma classe de RNAs conhecida como teoria da ressonância adaptativa ou *Adaptive Resonance Theory* (ART).

Hopfield (1982) utilizou a ideia de funções energia para formular um novo modo de se entender a computação executada por redes recorrentes com conexões sinápticas e simétricas e estabeleceu conexões com a física estatística (NASCIMENTO JÚNIOR e YONEYAMA, 2004; PACHECO e VELLASCO, 2007; SILVA *et al.*, 2010).

Outro desenvolvimento importante, em 1982, foi a formulação de mapas auto-organizáveis de Kohonen (1982), utilizando uma estrutura de rede unidimensional ou bidimensional, baseadas no mapa topológico presente no córtex cerebral (KOHONEN, 1982).

2.4.1 Características da Rede SOM

A rede SOM (*Self-Organizing Map*), ou Mapa Auto-Organizável de Kohonen, foi desenvolvida pelo finlandês Teuvo Kohonen em 1984, pesquisador da Universidade de Helsinki. Em vários trabalhos, esta rede é chamada de rede de Kohonen, o que pode provocar certa confusão, uma vez que Kohonen também propôs outros modelos de RNAs (CARVALHO, 1991; SASSI, 2006; LUDWIG JUNIOR e MONTGOMERY, 2007). Neste trabalho serão chamadas simplesmente de rede SOM.

A rede SOM é um tipo de RNA baseada em aprendizado não-supervisionado, sendo capaz de mapear um conjunto de dados de um espaço de entrada multidimensional, num conjunto finito de neurônios organizados em um arranjo normalmente unidimensional ou bidimensional (VESANTO e ALHONIEMI, 2000).

A rede SOM é um modelo de RNA competitivas em maior uso na atualidade e tem sido utilizada em uma gama de aplicações, incluindo agrupamento e visualização de dados multidimensionais (KOHONEN, 2001; COSTA e ANDRADE NETTO, 2007).

Algumas das características importantes da rede SOM são a compressão de informação e a tentativa de manutenção da topologia dos dados dos padrões de entrada, no qual as localizações espaciais dos neurônios da grade n-dimensional de saída são indicativas das características estatísticas intrínsecas, contidas nos padrões de entrada, que justificam o nome mapas auto-organizáveis.

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.

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 2.8 exhibe um mapa com arranjo hexagonal.

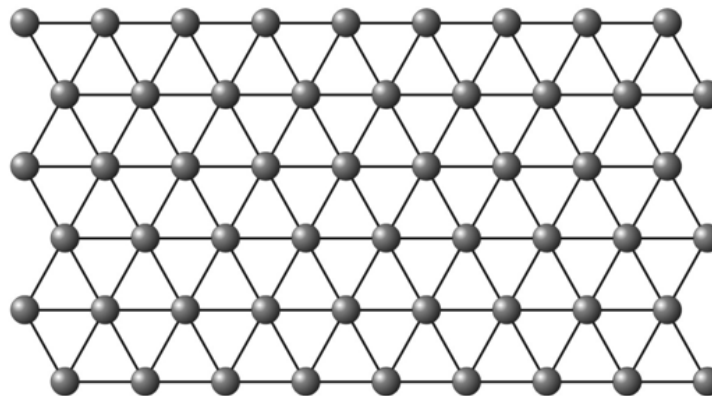


Figura 2.8 – Mapa com arranjo hexagonal para uma rede SOM com grade bidimensional.

Fonte: Adaptado de Kohonen (2001); Haykin (2001)

O formato do arranjo influencia diretamente na adaptação da rede SOM (HAYKIN, 2001; KOHONEN, 2001).

2.4.2 O Algoritmo de Aprendizado da Rede SOM

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

a) Competição. 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).

O princípio para o processo de aprendizado competitivo é a concorrência entre os neurônios, com o objetivo de determinar o vencedor da competição, pois o processo é não supervisionado, ou seja, não tem a saída desejada (SILVA *et al.*, 2010).

O prêmio para o neurônio vencedor da competição é o ajuste dos seus pesos, proporcionalmente aos valores do padrão de entrada apresentado, visando melhorar o seu estado para a próxima competição, se todas as conexões laterais deste neurônio vencedor forem nulas, implica que somente os seus pesos serão ajustados e assume-se a estratégia do vencedor leva tudo. Por causa disso, este tipo de rede é também chamado de rede do tipo o vencedor leva tudo ou *winner-takes-all* (SILVA *et al.*, 2010; BRAGA *et al.*, 2011).

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.

c) Adaptação Sináptica. Este último mecanismo permite que os neurônios excitados aumentem seus valores individuais da função discriminante em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos. Os ajustes feitos são tais que a resposta do neurônio vencedor à aplicação subsequente de um padrão de entrada similar é melhorada (HAYKIN, 2001).

Para cada dado apresentado à rede haverá uma competição entre todos os neurônios pelo direito de representá-lo. Vence a competição o neurônio que tiver o vetor de pesos mais próximo do vetor de dados.

O algoritmo da rede SOM utiliza o chamado aprendizado competitivo. Uma RNA com aprendizado competitivo é uma rede de uma única ou duas camadas (uni ou bidimensional) em que todos os neurônios recebem a mesma entrada. Cada neurônio computa o seu nível de ativação multiplicando o seu vetor de pesos pelo vetor de dados da entrada. O neurônio que tiver o maior nível de ativação é chamado de neurônio vencedor ou *Best Match Unit* (BMU), ou seja, o padrão de entrada que estiver sendo apresentado à rede provocará a ativação de apenas um neurônio de saída ou um neurônio por grupo (KOHONEN, 2001).

2.4.3 O Processo Competitivo da rede SOM

Para descrever os passos envolvidos com o processo de aprendizado competitivo utilizado na rede SOM, considera-se uma estrutura neural constituída de apenas uma camada neural (SILVA *et al.*, 2010). A Figura 2.9 mostra um modelo da rede SOM.

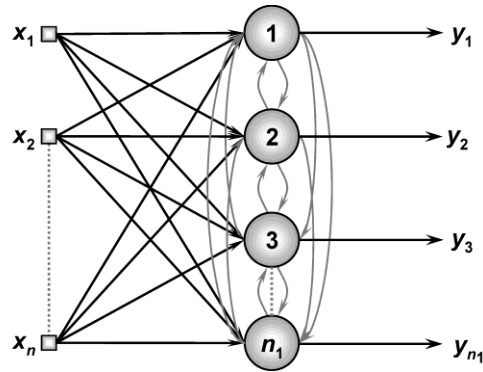


Figura 2.9 – Modelo da rede SOM. Fonte: Silva *et al.* (2010)

As conexões laterais entre os neurônios da Figura 2.9 assumem o papel de que um neurônio pode influenciar na resposta de saída produzida por outro neurônio.

Conforme Haykin (2001), se m representa a dimensão do espaço de entrada de dados, considere que um padrão vetor de entrada selecionado aleatoriamente do espaço de entrada seja representado pela Equação 2.4.

$$x = [x_1, x_2, \dots, x_m]^T \quad (2.4)$$

O vetor de peso sináptico de cada neurônio da grade tem a mesma dimensão que o espaço de entrada. Suponha que o vetor de peso sináptico do neurônio j seja representado pela Equação 2.5,

$$w_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, j = 1, 2, \dots, l \quad (2.5)$$

onde l é o número total de neurônios na grade. Para encontrar o melhor casamento do vetor de entrada x com os vetores de pesos sinápticos w_j , compare os produtos internos $w_j^T x$ para $j=1, 2, \dots, l$ e selecione o maior. Isto assume que o mesmo limiar seja aplicado a todos os neurônios; o limiar é o negativo das *bias*.

Assim, selecionando o neurônio com o maior produto interno $w_j^T x$, teremos de fato determinado a localização onde a vizinhança topológica dos neurônios excitados deve ser centrada (HAYKIN, 2001; LUDWIG JUNIOR e MONTGOMERY, 2007).

O melhor casamento, baseado na maximização do produto interno $w_j^T x$, é matematicamente equivalente a minimizar a distância euclidiana entre os vetores x e w_j . Se usarmos o índice $i(x)$ para identificar o neurônio que melhor casa com o vetor de entrada x , pode-se então determinar $i(x)$ aplicando a condição mostrada na Equação 2.6, que resume a essência do processo competitivo entre os neurônios.

$$i(x) = \arg \min_j \|x - w_j\|, j = 1, 2, \dots, l \quad (2.6)$$

Segundo Haykin (2001), dependendo da aplicação de interesse, a resposta da grade pode ser tanto o índice do neurônio vencedor, como o vetor de peso sináptico que está mais próximo do vetor de entrada em um sentido euclidiano.

O critério para escolha do neurônio vencedor pode-se basear em qualquer métrica que informe a distância entre o vetor de pesos dos neurônios. É comum o uso da distância euclidiana, o neurônio que apresenta a menor distância euclidiana entre o vetor de pesos é declarado o neurônio vencedor (LUDWIG JUNIOR e MONTGOMERY, 2007).

Se os vetores de entrada e de pesos já estiverem normalizados de forma unitária, o processo de aprendizagem obtém maior eficiência. A normalização unitária é efetuada bastando-se dividir cada vetor pelo seu respectivo módulo (SILVA *et al.*, 2010).

A Figura 2.10 mostra a regra de ajuste de pesos para o neurônio vencedor do processo competitivo, em que três neurônios representados por: $w^{(1)}$, $w^{(2)}$, $w^{(3)}$ que disputam a competição frente à apresentação do padrão $x^{(k)}$, sendo o mesmo composto por duas entradas x_1 e x_2 .

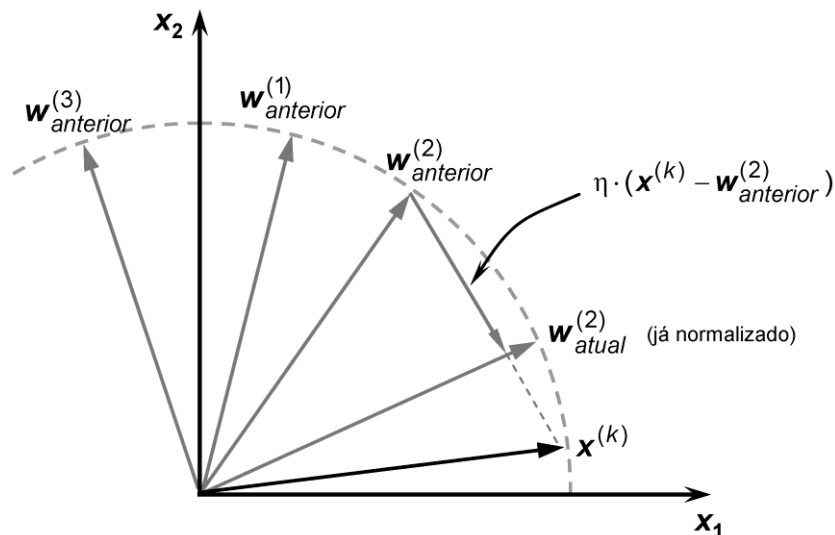


Figura 2.10 – Processo de ajuste do vetor de pesos do neurônio vencedor. Fonte: Silva *et al.* (2010)

A partir da Figura 2.10 observa-se que todos os vetores se encontram normalizados, o vetor $w^{(2)}$, é o vencedor da competição com os vetores $w^{(1)}$ e $w^{(3)}$, pois é o que está mais próximo do padrão $x^{(k)}$. Verifica-se que o ajuste consistiu tão somente em rotacionar o neurônio vencedor em direção ao vetor representando a amostra (SILVA *et al.*, 2010).

O neurônio vencedor é aquele que tiver o maior nível de ativação u_i , como o nível de ativação de um neurônio i é o produto entre o vetor de entrada $x^{(k)}$ e o vetor de pesos do neurônio $w^{(i)}$, conforme a Equação 2.7.

$$u_i = w^{(i)} x^{(k)} \quad (2.7)$$

O critério para a escolha do neurônio vencedor é o de similaridade entre $x^{(k)}$ e $w^{(i)}$. Este produto pode ser escrito conforme a Equação 2.8,

$$u_i = |w^{(i)}| |x^{(k)}| \cos \theta \quad (2.8)$$

onde θ é o ângulo entre os vetores $x^{(k)}$ e $w^{(i)}$.

A Figura 2.11 ilustra quatro classes disponíveis, em que se observa a distribuição dos vetores de pesos, após a estabilização da rede. Cada um destes vetores é representativo das amostras do problema, é constituído por dois sinais de entrada x_1 e x_2 .

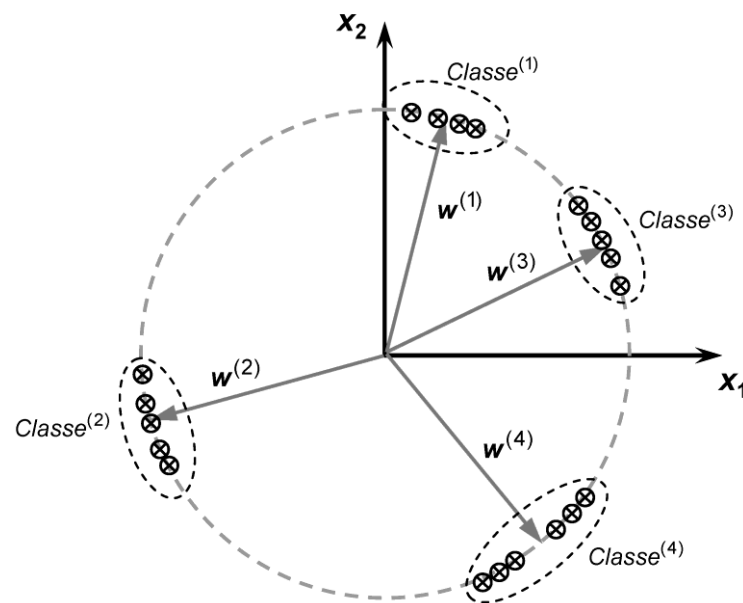


Figura 2.11 – Distribuição em espaço bidimensional dos vetores. Fonte: Silva *et al.* (2010)

Percebe-se na Figura 2.11, assim como na Figura 2.10, que todos os vetores acabam localizados no círculo unitário, à medida que são bidimensionais e estão unitariamente normalizados. Percebe-se, ainda, que os quatro vetores de pesos se posicionam nos centros dos *clusters* que representam as amostras (\otimes).

Desta forma, toma-se conhecimento que a estrutura neural representada nesta figura foi constituída por quatro neurônios, pois há somente quatro vetores de pesos $w^{(1)}$, $w^{(2)}$, $w^{(3)}$, $w^{(4)}$. (SILVA *et al.*, 2010).

Constata-se ainda na Figura 2.11 que, se a estrutura neural fosse constituída por cinco neurônios, ao invés de quatro, as amostras da classe⁽⁴⁾ seriam divididas em duas classes, pois a mesma classe está agrupando dois subconjuntos, embora muito próximos, aparecem em apenas um *cluster*.

2.4.4 O Processo Cooperativo da rede SOM

No processo cooperativo o neurônio vencedor localiza o centro de uma vizinhança topológica de neurônios cooperativos.

Segundo Haykin (2001), há evidência neurobiológica para a interação lateral entre um conjunto de neurônios excitados. Em particular um neurônio que está disparando tende a excitar mais fortemente os neurônios na sua vizinhança imediata que aqueles distantes dele, o que é intuitivamente razoável. Isso faz com que a vizinhança topológica em torno do neurônio vencedor decaia suavemente com a distância lateral.

No algoritmo original criado por Kohonen (1982), a vizinhança topológica é assumida tendo-se uma amplitude constante.

Observa-se que $h_{j,i}$ representa a vizinhança topológica centrada no neurônio vencedor i e que contenha um conjunto de neurônios excitados (cooperativos), sendo um neurônio típico deste conjunto representado por j . Considere que $d_{i,j}$ represente a distância lateral entre o neurônio vencedor i e o neurônio excitado j . Então pode-se assumir que a vizinhança topológica $h_{j,i}$ é uma função unimodal da distância $d_{j,i}$, desde que ela satisfaça duas exigências distintas:

- a) A vizinhança topológica $h_{j,i}$ é simétrica em relação ao ponto máximo definido por $d_{i,j} = 0$; em outras palavras, ela alcança o seu valor máximo no neurônio vencedor i para o qual a distância de $d_{j,i}$ é zero.
- b) A amplitude da vizinhança topológica $h_{j,i}$ decresce monotonamente com o aumento da distância lateral $d_{j,i}$, decaindo a zero para $d_{j,i} \longrightarrow \infty$; esta é condição necessária para a convergência (HAYKIN, 2001; KOHONEN, 2001).

A função que atende a estas exigências é a função gaussiana apresentada na Equação 2.9, sendo comum o seu uso.

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (2.9)$$

O parâmetro σ é a largura efetiva da vizinhança topológica como ilustrado na Figura 2.12.

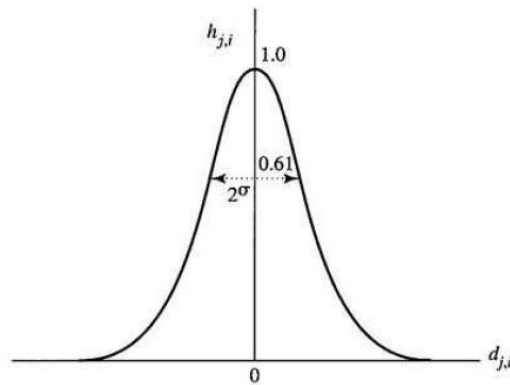


Figura 2.12 – Função de vizinhança gaussiana. Fonte: Haykin (2001)

A função de vizinhança topológica gaussiana faz com que o algoritmo da rede SOM convirja com maior rapidez (HAYKIN, 2001).

A função de vizinhança topológica mede o grau com o qual os neurônios excitados na vizinhança do neurônio vencedor participam do processo de aprendizagem.

No sentido qualitativo, a vizinhança topológica hexagonal oferece tradicionalmente resultados melhores que a vizinhança retangular (HAYKIN, 2001).

A Figura 2.13 mostra dois exemplos de vizinhança topológica, (a) vizinhança retangular e (b) vizinhança hexagonal sendo ($t_1 < t_2 < t_3$). No centro da vizinhança topológica observa-se o neurônio vencedor representado pela letra C.

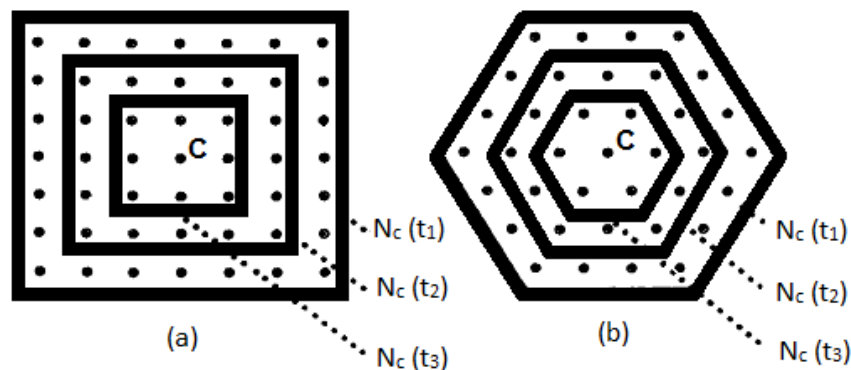


Figura 2.13 – Vizinhança topológica. Fonte: Kohonen (2001)

Para que a cooperação entre os neurônios vizinhos se mantenha, é necessário que a vizinhança topológica $h_{j,i}$ seja dependente da distância lateral $d_{j,i}$ entre o neurônio vencedor i e o neurônio excitado j no espaço de saída em vez de ser dependente de alguma medida de distância no espaço de entrada original, conforme verifica-se na Equação 5.7.

Na grade unidimensional, $d_{j,i}$ é um inteiro igual a $|j-i|$, e na grade bidimensional é definida pela Equação 2.10,

$$d_{j,i}^2 = \|r_j - r_i\|^2 \quad (2.10)$$

em que o vetor discreto r_j define a posição do neurônio excitado j e r_i define a posição discreta do neurônio vencedor i , sendo ambos medidos no espaço de saída discreto (HAYKIN, 2001).

Outra característica única do algoritmo SOM é que o tamanho da vizinhança topológica diminui com o tempo. Esta exigência é satisfeita fazendo-se com que a largura σ da função de vizinhança topológica $h_{j,i}$ diminua com o tempo.

Uma escolha típica para a dependência de σ com o tempo discreto n é o decaimento exponencial conforme mostra a Equação 2.11,

$$\sigma(x) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), n = 0,1,2,\dots, \quad (2.11)$$

onde σ_0 é o valor de σ na inicialização do algoritmo SOM, e τ_1 é a uma constante de tempo. Conseqüentemente, a vizinhança topológica assume uma forma variável no tempo, como mostra a Equação 2.12,

$$h_{j,i(x)}(n) = \exp\left(\frac{d_{j,i}^2}{2\sigma^2(n)}\right), n = 0,1,2,\dots, \quad (2.12)$$

onde $\sigma(x)$ é definido pela Equação 2.11. Assim, quando o número de iterações n aumenta, a largura $\sigma(x)$ decresce a uma taxa exponencial e a vizinhança topológica diminui de maneira correspondente.

Outra forma de verificar a variação da função de vizinhança $h_{j,i(x)}(n)$ em torno de um neurônio vencedor $i(x)$ é o propósito de um $h_{j,i(x)}(n)$ largo, é essencialmente correlacionar as direções das atualizações dos pesos de um grande número de neurônios excitados da grade.

Quando a largura de $h_{j,i(x)}(n)$ é diminuída, também diminui o número de neurônios cujas direções de atualização são correlacionadas.

Segundo Haykin (2001), é um desperdício de recursos computacionais mover um grande número de graus de liberdade em torno de um neurônio vencedor de forma correlacionada, como no caso do algoritmo SOM padrão.

É melhor utilizar uma forma de treinamento da SOM normalizada, na qual usa-se um número muito menor de graus de liberdade normalizados, esta operação é facilmente realizada na forma discreta tendo uma função de vizinhança $h_{j,i(x)}(n)$ de largura constante, mas aumentando gradualmente o número total de neurônios (HAYKIN, 2001).

Os novos neurônios são inseridos na metade da distância entre os neurônios antigos, e a suavidade do algoritmo SOM garante que os novos neurônios se insiram na adaptação sináptica de maneira mais suave.

2.4.5 O Processo Adaptativo Sináptico da rede SOM

O último processo envolvido na formação do mapa auto-organizável é o processo adaptativo sináptico.

Para que a grade seja auto-organizável, é necessário que o vetor de peso sináptico w_j do neurônio j da grade se modifique em relação ao vetor de entrada x .

No postulado de aprendizagem de Hebb (1949), um peso sináptico é aumentado com uma ocorrência simultânea de atividades pré-sináptica e pós-sináptica, o uso de tal regra é muito adequado para aprendizagem associativa. Entretanto, para o tipo de aprendizagem não-supervisionada, considerado a hipótese Hebbiana na sua forma básica, não é satisfatória pelas seguintes razões: as modificações das conectividades ocorrem apenas em uma direção, o que leva no final todos os pesos à saturação (HAYKIN, 2001).

Para superar este problema, modifica-se a hipótese Hebbiana, incluindo um termo de esquecimento – $g(y_j)w_j$ onde w_j é o vetor peso sináptico do neurônio j e $g(y_j)$ é uma função escalar positiva da resposta y_j . A única exigência imposta à função $g(y_j)$ é que o termo constante da expansão de $g(y_j)$ seja zero, de modo que se pode escrever de acordo a Equação 2.13.

$$g(y_j) = 0 \quad \text{para} \quad y_j = 0 \quad (2.13)$$

Segundo Haykin (2001), o significado desta exigência será explicado a seguir. Dada esta função, pode-se então expressar a modificação do vetor peso do neurônio j da grade como mostra a Equação 2.14,

$$\Delta w_j = \eta y_j x - g(y) w_j \quad (2.14)$$

onde η é o parâmetro da taxa de aprendizagem do algoritmo, o primeiro termo do lado direito da Equação 2.14 é o termo Hebbiano e o segundo termo é o termo de esquecimento.

Para atender a exigência da Equação 2.13, será utilizada uma função linear para $g(y_j)$, como mostra a Equação 2.15.

$$g(y_j) = \eta y_j \quad (2.15)$$

É possível simplificar mais a Equação 2.15 conforme a equação abaixo:

$$y_j = h_{j,i(x)} \quad (2.16)$$

Utilizando as Equações 2.15 e 2.16 na Equação 2.14 chega-se a:

$$\Delta w_j = \eta^h_{j,i(x)} (x - w_j) \quad (2.17)$$

Por fim, utilizando o formalismo de tempo discreto, dado o vetor de peso sináptico $w_j(n)$ do neurônio j no tempo n , o vetor de peso atualizado $w_j(n+1)$ no tempo $n+1$ é definido por Kohonen (2001), conforme a Equação 2.18,

$$w_j(n+1) = w_j(n) + \eta(n) h_{j,i(x)}(n) (x - w_j(n)) \quad (2.18)$$

A Equação 2.18 é aplicada a todos os neurônios da grade que se encontram dentro da vizinhança topológica do neurônio vencedor i , e tem o efeito de mover o vetor peso sináptico w_i do neurônio i em direção ao vetor de entrada x .

Para Haykin (2001), o parâmetro da taxa de aprendizagem $\eta(n)$ deve ser variável no tempo como mostrado na Equação 2.18 anteriormente, que corresponde ao caso da aproximação aleatória.

A taxa de aprendizagem deve começar com um valor inicial η_0 e então descrever gradualmente com o aumento do tempo n , esta exigência pode ser satisfeita escolhendo-se um decaimento exponencial para $\eta(n)$, como mostra a Equação 2.19,

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_2}\right), n = 0, 1, 2, \dots, \quad (2.19)$$

onde τ_2 é uma outra constante de tempo do algoritmo SOM. Apesar das fórmulas de decaimento exponencial descritas nas Equações 2.11 e 2.19 para largura da função de vizinhança e o parâmetro da taxa de aprendizagem, respectivamente, poderem não ser ótimas, elas são normalmente adequadas para a formação do mapa de características de uma maneira auto-organizada (HAYKIN, 2001).

2.4.6 As duas Fases do Processo Adaptativo da rede SOM

Num estado inicial de desordem completa, é surpreendente como o algoritmo SOM gradualmente leva a uma representação organizada de padrões de ativação retirados do espaço de entrada, desde que os parâmetros do algoritmo sejam selecionados adequadamente (HAYKIN, 2001).

De acordo com Kohonen (2001), é possível decompor a adaptação dos pesos sinápticos da grade, calculada em duas fases: uma fase de ordenação ou de auto-organização seguida por uma fase de convergência, estas duas fases do processo adaptativo da rede SOM são descritas a seguir:

- a) Fase de auto-organização ou de ordenação. É durante esta fase do processo adaptativo que ocorre a ordenação topológica dos vetores de peso.

Segundo Haykin (2001) a fase de ordenação pode exigir mil iterações do algoritmo SOM, em geral mais. Nesta fase deve-se ter cautela na escolha do parâmetro de aprendizagem e da função de vizinhança.

- b) Fase de convergência. Esta fase do processo adaptativo é necessária para realizar uma sintonia fina do mapa de características e assim produzir uma quantização estatística precisa do espaço de entrada.

O número de iterações que constituem a fase de convergência deve ser no mínimo quinhentas vezes o número de neurônios na grade. Assim, a fase de convergência pode durar milhares ou dezenas de milhares de iterações.

2.4.7 Medidas de Qualidade da rede SOM

O Erro de Quantização (EQ) corresponde à média das distâncias entre cada vetor de dados x e o correspondente vetor de pesos do neurônio BMU (*Best Match Unit*). A medida corresponde à resolução do mapa e está apresentada na Equação 2.20 (KIVILUOTO, 1995).

$$EQ = \frac{1}{N} \sum_{k=1}^N \|x - w_i\| \quad (2.20)$$

Estima-se que quanto menor o EQ, mais bem ajustado o neurônio vencedor estará aos vetores de entrada (FARIA *et al.*, 2010).

O Erro Topográfico (ET) quantifica a capacidade do mapa em representar a topologia dos dados de entrada (KIVILUOTO, 1995). Para cada vetor de dados de entrada x são calculados seu primeiro BMU e o seu segundo BMU e toda vez que eles não forem adjacentes (vizinhos), aumenta-se o erro em uma unidade, tirando-se depois a média pelo número total

de vetores, onde $u(x_k)$ igual a 1 se o primeiro e o segundo BMU não são adjacentes, caso contrário igual a 0 (FARIA *et al.*, 2010).

A seguir a Equação 2.21 que calcula o erro topográfico.

$$ET = \frac{1}{N} \sum_{k=1}^N u(x_k) \quad (2.21)$$

Quando os valores do EQ e ET são muito baixos, suspeita-se de *overfitting*. A rede SOM, na tentativa de representar o mais fielmente possível os dados, dobra-se de forma que acaba representando exatamente os dados, podendo perder a sua capacidade de generalização.

O fenômeno inverso, o *underfitting*, ocorre quando um mapa é rígido demais. Isso pode acontecer quando há poucos neurônios para representar um número proporcionalmente grande de dados ou o raio de vizinhança final da função gaussiana for maior que 1 durante o treinamento. Neste caso, os valores de EQ podem ser mais altos (os vetores de pesos dos neurônios encontram-se, em média, menos próximos dos vetores de dados) (KASKI e KOHONEN, 1997; SVENSÉN, 1998).

2.5 MULTILAYER PERCEPTRON

Uma RNA do tipo *Multilayer Perceptron* (MLP) consiste de um conjunto de unidades (nós ou neurônios), que constituem a camada de entrada, uma ou mais camadas ocultas e uma camada de saída, onde o sinal de entrada se propaga pela RNA camada por camada. Na Figura 2.14 é apresentada a estrutura básica de uma RNA do tipo MLP.

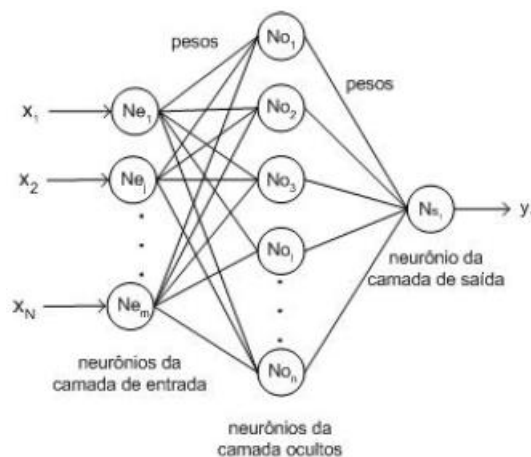


Figura 2.14 – Estrutura básica da RNA do tipo MLP. Fonte: Sassi (2006)

Na Figura 2.14 podem-se observar os dados (vetores de dados) de entrada da rede (X_1, \dots, X_N), os neurônios da camada de entrada da rede (N_{e1}, \dots, N_{em}) com seus respectivos pesos, os

neurônios que formam a camada intermediária da rede (N_{o1}, \dots, N_{on}) e a camada de saída (N_{s1}), formada por um neurônio.

As RNAs possuem a capacidade de aprender por exemplos e fazer interpolações e extrapolações do que aprenderam. Um conjunto de procedimentos bem definidos, para adaptar os pesos de uma RNA para que ela possa aprender uma determinada função, é chamado algoritmo de treinamento ou de aprendizado (BIGUS, 1996).

O aprendizado de uma RNA usa um conjunto de dados correspondentes a uma amostra de sinais para entrada e saída do sistema. Para esse treinamento, a rede utiliza algoritmos de aprendizado.

Inicialmente a rede permanece inerte e o algoritmo de aprendizado modifica individualmente os pesos das interconexões de tal forma que o comportamento da rede reflita a ação desejada. Em outras palavras, a rede pode alterar sua estrutura interna de maneira incremental até que se alcance o desempenho esperado de estimação dos dados (SIMÕES e SHAW, 2007).

O algoritmo de retroprogramação do erro (*error back-propagation*), utilizado na MLP, consiste, basicamente, em determinar as variações nos pesos sinápticos da RNA, tendo como objetivo minimizar o erro obtido na saída através do aprendizado do vetor de treinamento (entrada-saída) (RUMELHART *et al.*, 1986).

Para isso, o algoritmo baseia-se no método do gradiente descendente, o qual, dada uma medida do erro, procura modificar o conjunto de pesos w_{ij} da rede, reduzindo o erro na direção mais íngreme da superfície definida no espaço w (HAYKIN, 2001).

Em resumo, o algoritmo do gradiente descendente estabelece mudanças nos pesos w_{ij} por uma quantidade $\Delta_{w_{ij}}$ proporcional ao gradiente do erro.

O algoritmo de retroprogramação do erro (*error back-propagation*) funciona da seguinte forma: apresenta-se um padrão à camada de entrada da rede, esse padrão é processado camada por camada até que a camada de saída forneça a resposta processada, f_{MLP} , como mostra a Equação (2.22):

$$f_{MLP}(x) = \varphi \left(\sum_1^{Non} v_l \cdot \varphi \left(\sum w_{lv} x_l + b_{l0} \right) + b_0 \right) \quad (2.22)$$

onde v_l e w_{lv} são pesos sinápticos; b_{l0} e b_0 são os biases; e φ a função de ativação, comumente especificada como sendo a função sigmóide, conforme Figura 2.15.

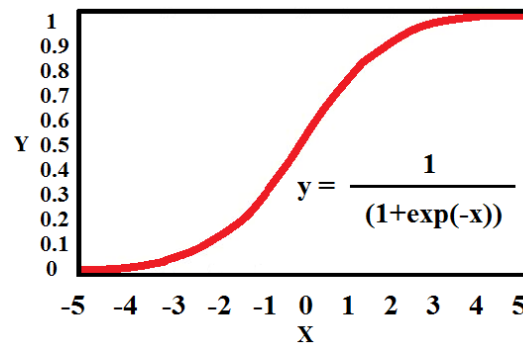


Figura 2.15 – Função sigmóide. Fonte: Adaptado de Sassi, Silva e Hernandez (2008)

O princípio deste algoritmo, para cálculo dos erros nas camadas intermediárias, é o seu cálculo por retroalimentação, possibilitando, desta forma, o ajuste dos pesos proporcionalmente aos valores das conexões entre camadas.

De acordo com Haykin (2001), a MLP possui as seguintes características: função de ativação não-linear (sigmoidal), uma ou mais camadas de neurônios ocultos e um alto grau de conectividade. Uma RNA MLP treinada com o algoritmo de retropropagação realiza um mapeamento não-linear de entrada-saída.

Ainda de acordo com Haykin (2001), matematicamente, uma RNA é um método para determinar uma função $f: A \rightarrow B$ onde A é um conjunto de dados, sendo que alguns elementos de B são conhecidos *a priori*.

2.5.1 Critérios de Parada de Treinamento da MLP

De acordo com Simões e Shaw (2007), em um processo iterativo como o *algoritmo error back-propagation*, uma questão importante é sobre o critério de parada do treinamento.

Em termos absolutos, a solução final ocorrerá para o índice de desempenho (erro global) igual a zero ou dentro de um valor muito pequeno. Contudo, se durante a fase de treinamento o erro não for pequeno e não tiver tendência a diminuir, dois enfoques podem ser adotados:

a) Limitar o número de iterações, ou seja, o treinamento cessa após ter ocorrido um valor limite, prefixado, de épocas de treinamento;

b) Amostrar e tirar a média de certo número de épocas, por exemplo, 500 épocas. Se o erro médio do último conjunto de 500 épocas não for melhor que o das 500 anteriores o treinamento deve ser cessado, indicando que um mínimo local foi alcançado. Depois disso, a rede deve estar ainda pronta para ser reutilizada para testes.

Se o desempenho não for satisfatório, aos pesos devem ser adicionados um pouco de ruído branco para auxiliar a rede a sair do mínimo local. Ou então, a rede deve ser

completamente reinicializada e o processo de treinamento deve ser retomado (SIMÕES e SHAW, 2007).

A MLP é normalmente utilizada para problemas como:

- Classificação: a MLP é treinada através de amostras de dados de cada grupo e instruções de como realizar a classificação, fazendo com que a rede aprenda as características próprias de cada grupo;

- Aproximação (ou análise de regressão): inclui previsão e modelagem de séries temporais;

- Outras áreas de aplicação incluem: sistema de identificação e controle (controle veicular, de processos, entre outros), tomada de decisão em jogos (gamão, xadrez, corridas), diagnósticos médicos, aplicações financeiras (sistemas de negociação automáticos), *Data Mining* (descoberta de conhecimento nas bases de dados, ou KDD), visualização e filtragem de spam de e-mail.

3 METODOLOGIA DA PESQUISA

Neste capítulo é apresentada a metodologia da pesquisa, a base de dados dos empregados da empresa de *Courier* utilizada nos experimentos, os parâmetros utilizados na estrutura da rede SOM e na estrutura da rede MLP, a plataforma de *hardware* e *software* e informações sobre a pesquisa do tipo *survey* aplicada nos empregados da empresa.

3.1 BASE DE DADOS EXPERIMENTAL

A metodologia adotada foi definida como exploratória e experimental.

A Pesquisa Exploratória têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses. Pode-se dizer que estas pesquisas têm como objetivo principal o aprimoramento de ideias ou a descoberta de intuições. Seu planejamento é, portanto, bastante flexível para que possibilite a consideração dos mais variados aspectos relativos ao fato estudado (GIL, 2002).

Na maioria dos casos, essas pesquisas envolvem: (a) levantamento bibliográfico; (b) entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado e (c) análise de exemplos que "estimulem a compreensão". Embora o planejamento da pesquisa exploratória seja bastante flexível, na maioria dos casos assume a forma de pesquisa bibliográfica ou de estudo de caso (GIL, 2002).

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 e presenteísmo, descoberta de conhecimento em bases de dados, mineração de dados, redes neurais artificiais, rede SOM e rede MLP.

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 na área operacional dessa empresa de *Courier*.

As Empresas de *Courier* são empresas que tem como atividade principal a prestação de serviços de transporte nacional e internacional expresso, porta a porta, em pelo menos três continentes de remessa expressa destinada a terceiros, em fluxo regular e contínuo, tanto na importação como na-exportação (DOU, 1985).

Os motivos do absenteísmo com CID foram estratificados em 21 categorias conforme a Tabela 3.1 e os registros do absenteísmo sem CID foram estratificados em 7 categorias

conforme mostra a Tabela 3.2, 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.

Observa-se na Tabela 3.1 vinte e uma categorias do CID, que 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).

Tabela 3.1 – Lista de categorias com (CID)

1	Algumas doenças infecciosas e parasitárias (A00-B99)
2	Neoplasias [tumores] (C00-D48)
3	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários (D50-D89)
4	Doenças endócrinas, nutricionais e metabólicas (E00-E90)
5	Transtornos mentais e comportamentais (F00-F99)
6	Doenças do sistema nervoso (G00-G99)
7	Doenças do olho e anexos (H00-H59)
8	Doenças do ouvido e da apófise mastóide (H60-H95)
9	Doenças do aparelho circulatório (I00-I99)
10	Doenças do aparelho respiratório (J00-J99)
11	Doenças do aparelho digestivo (K00-K93)
12	Doenças da pele e do tecido subcutâneo (L00-L99)
13	Doenças do sistema osteomuscular e do tecido conjuntivo (M00-M99)
14	Doenças do aparelho geniturinário (N00-N99)
15	Gravidez, parto e puerpério (O00-O99)
16	Algumas afecções originadas no período perinatal (P00-P96)
17	Malformações congênitas, deformidades e anomalias cromossômicas (Q00-Q99)
18	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte (R00-R99)
19	Lesões, envenenamento e algumas outras conseqüências de causas externas (S00-T98)
20	Causas externas de morbidade e de mortalidade (V01-Y98)
21	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde (Z00-Z99)

Fonte: Adaptado de Datasus (2000)

A Tabela 3.2 mostra as 7 categorias de absenteísmo não classificados pelo CID.

Tabela 3.2 – Lista de categorias sem (CID)

22	Acompanhante
23	Consulta médica
24	Doação de sangue
25	Exame
26	Falta injustificada
27	Fisioterapia
28	Consulta odontológica

A categoria CID que se destaca no presenteísmo é a categoria de número 6 - Doenças do sistema nervoso (G00-G99).

A Tabela 3.3 mostra parte da base de dados do absenteísmo na empresa de *Courier*.

Código_Nome	Motivo da ausencia	Estacao do ano	Transporte R\$	Dist. da Res para o Trab.	Tempo de servico	Idade	Carga de trabalho	Meta Atingida	Falta Disciplinar	Escolaridade	Numero de Filhos	Etilista	Tabagista	Animas de Est.	Peso	Altura	IMC	T. arast. horas	Mês	Dia da Semana
11	26	1	289,08	36,4	13	33	239,554	96,62	0	1	2	1	0	1	90	1,72	30,4	4	7	3
3	23	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	2	7	4
7	7	1	279,4	4,9	14	39	239,554	96,62	0	1	2	1	1	0	68	1,68	24,1	4	7	5
11	23	1	289,08	36,4	13	33	239,554	96,62	0	1	2	1	0	1	90	1,72	30,4	2	7	5
3	23	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	2	7	6
10	22	1	361,2	51,9	3	28	239,554	96,62	0	1	1	1	0	4	80	1,72	27	8	7	6
20	23	1	260	49,8	11	36	239,554	96,62	0	1	4	1	0	0	65	1,68	23	4	7	6
14	19	1	154,66	12,1	14	34	239,554	96,62	0	1	2	1	0	0	95	1,96	24,7	40	7	2
1	22	1	235,4	11,8	14	37	239,554	96,62	0	3	1	0	0	1	88	1,72	29,7	8	7	2
20	1	1	260	49,8	11	36	239,554	96,62	0	1	4	1	0	0	65	1,68	23	8	7	2
20	1	1	260	49,8	11	36	239,554	96,62	0	1	4	1	0	0	65	1,68	23	8	7	3
20	11	1	260	49,8	11	36	239,554	96,62	0	1	4	1	0	0	65	1,68	23	8	7	4
3	11	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	1	7	4
3	23	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	4	7	4
24	14	1	246,4	24,7	16	41	239,554	96,62	0	1	0	1	0	0	67	1,7	23,2	8	7	6
3	23	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	2	7	6
3	21	1	179,08	50,6	18	38	239,554	96,62	0	1	0	1	0	0	89	1,70	30,8	8	7	2
6	11	1	189,56	29,6	13	33	239,554	96,62	0	1	2	0	0	2	69	1,67	24,7	8	7	5
33	23	1	246,6	25,4	14	47	205,917	91,96	0	1	2	0	0	1	86	1,65	31,6	2	8	4
18	10	1	330,88	15,8	4	28	205,917	91,96	0	2	0	0	0	0	84	1,82	25,4	8	8	4
3	11	1	179,08	50,6	18	38	205,917	91,96	0	1	0	1	0	0	89	1,70	30,8	1	8	2
10	29	1	361,2	51,9	3	28	205,917	91,96	0	1	1	1	0	4	80	1,72	27	40	8	2
20	28	1	260	49,8	11	36	205,917	91,96	0	1	4	1	0	0	65	1,68	23	4	8	6
11	18	1	289,08	36,4	13	33	205,917	91,96	0	1	2	1	0	1	90	1,72	30,4	8	8	2
10	25	1	361,2	51,9	3	28	205,917	91,96	0	1	1	1	0	4	80	1,72	27	7	8	2
11	23	1	289,08	36,4	13	33	205,917	91,96	0	1	2	1	0	1	90	1,72	30,4	1	8	3
30	28	1	157,5	26,8	6	29	205,917	91,96	0	1	0	1	1	0	75	1,85	21,9	4	8	4
11	18	1	289,08	36,4	13	33	205,917	91,96	0	1	2	1	0	1	90	1,72	30,4	8	8	4
3	23	1	179,08	50,6	18	38	205,917	91,96	0	1	0	1	0	0	89	1,70	30,8	2	8	6
3	18	1	179,08	50,6	18	38	205,917	91,96	0	1	0	1	0	0	89	1,70	30,8	8	8	2
2	18	1	235,4	29,6	12	48	205,917	91,96	0	1	1	0	1	5	88	1,63	33,1	8	8	5
1	23	1	235,4	11,8	14	37	205,917	91,96	0	3	1	0	0	1	88	1,72	29,7	4	8	5
2	18	1	235,4	29,6	12	48	205,917	91,96	0	1	1	0	1	5	88	1,63	33,1	8	8	2
3	23	1	179,08	50,6	18	38	205,917	91,96	0	1	0	1	0	0	89	1,70	30,8	2	8	2
10	23	1	361,2	51,9	3	28	205,917	91,96	0	1	1	1	0	4	80	1,72	27	1	8	2
11	24	1	289,08	36,4	13	33	205,917	91,96	0	1	2	1	0	1	90	1,72	30,4	8	8	3

Tabela 3.3 – Base de Dados do absentismo

3.1.1 Realização dos Experimentos

A seguinte metodologia experimental foi adotada para a realização dos experimentos:

- a base de dados dos empregados foi preparada e submetida à rede SOM;
- a rede SOM processou os registros (informações dos empregados) da base de dados gerando os agrupamentos;
- os registros agrupados pela rede SOM foram classificados pela rede MLP a fim de validar os agrupamentos. Isto foi verificado com base no percentual de classificação correta dos empregados agrupados.

A Figura 3.1 ilustra a metodologia experimental do trabalho.

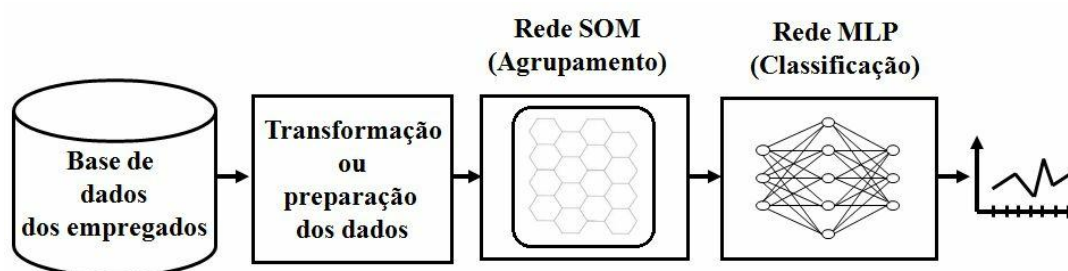


Figura 3.1 – Metodologia experimental. Fonte: Sassi (2012)

3.2 FERRAMENTAS UTILIZADAS E PLATAFORMAS DE ENSAIO

A plataforma de *hardware* utilizada nos experimentos foi um computador com processador Intel® Core™ 2 Duo CPU T6600 de 2,20 GHz com 4,00 GB de memória RAM DDR3, 320 GB de disco rígido e sistema operacional Windows 7 de 64 Bits.

Para a tabulação da base de dados, tanto para utilização na rede SOM quanto na pesquisa tipo *survey* foi utilizado o MS Excel.

Com relação ao *software* utilizado para gerar os Mapas Auto-Organizáveis, foi utilizado o *software* Viscovery® SOMine versão 5.2.2.

A escolha do *software* SOMine para utilização nos experimentos baseou-se inicialmente na bibliografia consultada, no caso o livro do criador dos Mapas Auto-Organizáveis, Teuvo Kohonen (KOHONEN, 2001).

Conforme Kohonen (2001), o Viscovery SOMine é um pacote comercial produzido por SOM Eudaptics GmbH (<http://www.eudaptics.co.at>). O *software* é considerado amigável, flexível e poderoso, especialmente em problemas estatísticos.

A escolha do *software* baseou-se ainda em trabalhos publicados que relatam a sua utilização (ZHUANG *et al.*, 2006; LOZANO *et al.*, 2010; CANTO *et al.*, 2011).

O *software* utilizado na rede MLP foi o *Multiple Back-Propagation* versão 2.2.2. A escolha do *software* baseou-se em trabalhos publicados que relatam a sua utilização (LOPES e RIBEIRO, 2010, 2009).

3.3 PARÂMETROS DA REDE SOM

Os Parâmetros utilizados na estrutura da rede SOM são os seguintes:

a) Dimensões: número de neurônios usados nos experimentos: $25 \times 25 = 625$.

Conforme Vesanto (2002), uma estimativa razoável para a escolha do número de neurônios é $n = 5\sqrt{D}$, onde D é a quantidade de dados de entrada. Sendo 740 registros multiplicados por 21 atributos chega-se a 15.540 dados, substituindo a equação $n = 5\sqrt{15540} = 623,29$.

b) Dimensões: número de neurônios usados na pesquisa tipo *survey*: $30 \times 30 = 900$.

Vale ressaltar que a dimensão da rede SOM, usada na pesquisa do tipo *survey*, é maior em função da necessidade de visualização dos *clusters*, apesar da pequena quantidade de dados (406).

c) Vizinhança topológica: hexagonal.

Segundo Haykin (2001), a vizinhança topológica hexagonal oferece tradicionalmente resultados melhores que a vizinhança retangular. Para Kohonen (2001), a vizinhança topológica hexagonal proporciona melhor qualidade na análise visual dos agrupamentos. Por esse motivo foi escolhida.

d) Função de vizinhança topológica: gaussiana.

A função de vizinhança topológica gaussiana faz com que o algoritmo da rede SOM convirja com maior rapidez (HAYKIN, 2001). Para Kohonem (2001), a função de vizinhança baseada em uma gaussiana tende a destacar melhor os agrupamentos.

Os Parâmetros de treinamento utilizados na rede SOM são os seguintes:

a) Número de épocas igual a 3000 (na *survey* igual a 1000).

b) Taxa de aprendizado igual a $\eta = 0,5$ na fase inicial de treinamento e, na fase de convergência, a taxa foi de $\eta = 0,05$ (KASKI e KOHONEN, 1997).

Mesmo com a utilização das heurísticas propostas por Kaski e Kohonen (1997), Haykin (2001), Kohonen (2001) e Vesanto (2002) para os parâmetros da rede SOM, é recomendado à realização de diversas configurações para teste da rede SOM, antes da escolha

definitiva dos parâmetros. Antes dos parâmetros adotados nos experimentos foram testados diversos parâmetros até a escolha definitiva.

As medidas de qualidade da rede SOM:

a) Erro de Quantização: 0,7712.

Estima-se que quanto menor o Erro de Quantização, mais bem ajustado o neurônio vencedor estará aos vetores de entrada (FARIA *et al.*, 2010).

b) Erro Topográfico: 0,087.

O Erro Topográfico (ET) quantifica a capacidade do mapa em representar a topologia dos dados de entrada (KIVILUOTO, 1995).

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

3.4 PARÂMETROS DA *MULTILAYER PERCEPTRON*

As redes MLP são caracterizadas por possuírem mais de uma camada oculta. Para este trabalho, a rede MLP, com duas camadas ocultas, apresentou melhor desempenho.

A fim de minimizar o erro da MLP, alguns parâmetros foram aperfeiçoados como, por exemplo, o número de neurônios das camadas intermediárias, o que permitiu a construção de uma rede mais eficiente.

A seguir são apresentados os parâmetros utilizados para construção da MLP: número de neurônios de entrada igual a 21, número de camadas igual a 2, número de neurônios nas camadas ocultas igual a 10, taxa inicial de aprendizagem igual a 0,5 com decaimento de 1% a cada 5 épocas, fator de momento inicial igual 0,9 com decaimento de 1% a cada 100 épocas, o critério de parada foi o número máximo de épocas igual a 1500. O conjunto de dados foi separado em 60% para a fase de treinamento e 40% para a fase de teste.

3.5 CARACTERÍSTICAS DA PESQUISA DO TIPO *SURVEY*

Segundo Gil (2002), a pesquisa do tipo *survey* permite realizar levantamento em condições controladas a partir de subconjuntos da população chamados amostras e selecionadas ao acaso ou em função de características particulares. São dados quantitativos passíveis de análise e de generalizações, que possa descrever grupos/pessoas (ou explicar) no

próprio ambiente. Esta amostra é controlada (o ambiente somado à amostra, que dá condições de generalização).

A pesquisa foi realizada por meio da aplicação de um questionário fechado (Apêndice A), contendo 14 questões adaptadas de Luft *et al.* (2007) para os 36 empregados que compõem a base de dados estudada.

Para Silva e Menezes (2001), perguntas fechadas são aquelas em que são oferecidas as respostas a serem escolhidas. Já as perguntas abertas são aquelas em que o sujeito tem a liberdade de dissertar o texto ao responder.

O objetivo da pesquisa foi verificar o nível de estresse percebido pelos empregados e as possíveis relações com o absenteísmo e o presenteísmo, já que este é considerado como o fator desencadeador dos dois fenômenos (FAVA, 2011).

Foi utilizada A Escala Likert, que é um tipo de escala de resposta psicométrica usada comumente em questionários, além de ser a escala mais usada em pesquisas de opinião. Esta escala tem seu nome devido à publicação de um relatório explicando seu uso por Rensis Likert (LIKERT, 1932). Ao responderem a um questionário baseado nesta escala, os perguntados especificam seu nível de concordância com uma afirmação.

A escala para as respostas do questionário são apresentadas a seguir: 0 = Nunca, 1 = Quase nunca, 2 = Às vezes, 3 = Quase sempre, 4 = Sempre.

Foram respondidos 29 questionários, o que correspondem a cerca de 80% dos empregados e é considerado um excelente resultado para este tipo de pesquisa (SALLES e IOZZI, 2010).

Os resultados obtidos com a pesquisa foram tabulados no MS EXCEL.

4 RESULTADOS E DISCUSSÕES

Neste capítulo é apresentada a aplicação da pesquisa do tipo *survey*, a realização dos experimentos com a rede SOM, com a rede MLP e também os respectivos resultados.

4.1 RESULTADOS DA PESQUISA DO TIPO *SURVEY*

O estresse, segundo Coleman (1992), Primo (2008), Laranjeira (2009) e Umann (2011) é responsável por grande parte da perda da produtividade no trabalho durante a vida do empregado.

Verificar, portanto, a presença de estresse nos empregados da empresa em questão, é fundamental para o estudo realizado neste trabalho, porque ele, o estresse, é considerado fator causador do absenteísmo e presenteísmo. Isto justifica a aplicação desta pesquisa tipo *survey*.

As respostas dadas pelos 29 funcionários às 14 questões do questionário aplicado são apresentadas nos mapas gerados pela rede SOM, ilustradas nas Figuras ímpares de 4.1 até 4.27. A distribuição relativa das respostas é apresentada nas Figuras pares de 4.2 até 4.28.

A Figura 4.1 mostra os *clusters* gerados pela rede SOM para a primeira questão: “Você tem ficado triste por causa de algo que aconteceu inesperadamente?”.

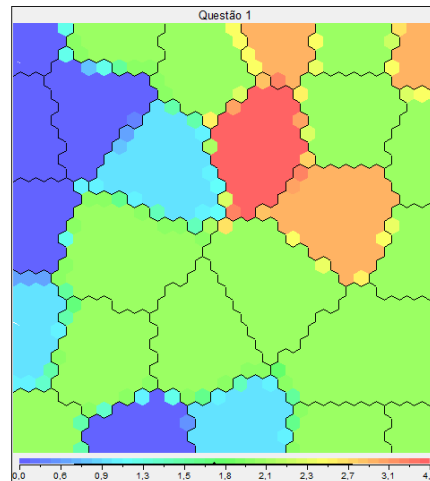


Figura 4.1 – *Clusters* (Questão 1)

A média aferida nessa questão foi de 1,72.

A Figura 4.2 mostra a distribuição relativa das respostas. Na primeira questão, a maior concentração das respostas está em “Às vezes” com 59% e menor concentração em “Sempre” com 4%. Observa-se que entre os respondentes 14% tem ficado triste e 27% não tem ficado triste.

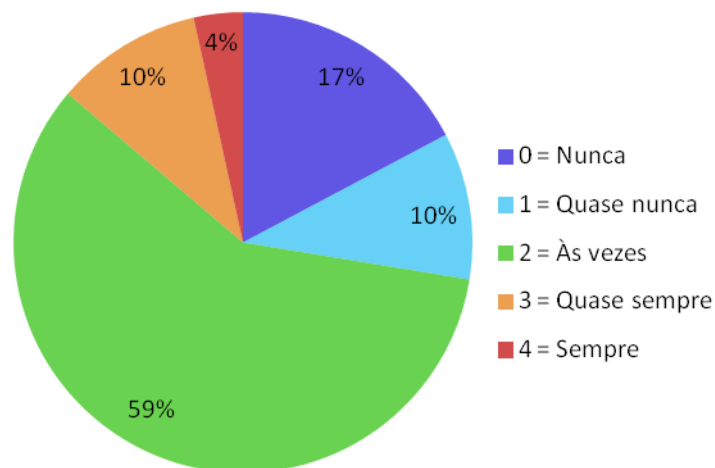


Figura 4.2 – Distribuição relativa (Questão 1)

A Figura 4.3 mostra os *clusters* gerados pela rede SOM para a segunda questão: “Você tem se sentido incapaz de controlar as coisas importantes em sua vida?”.

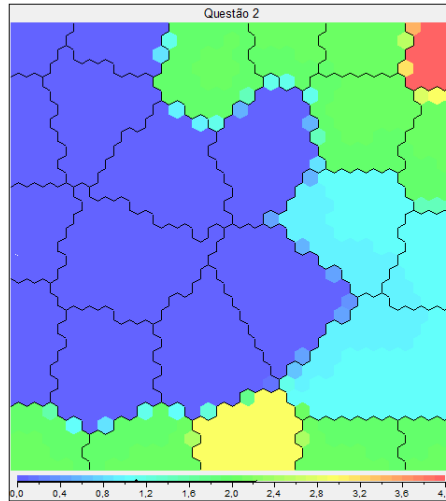


Figura 4.3 – *Clusters* (Questão 2)

A média aferida nessa questão foi de 1,14.

A Figura 4.4 mostra a distribuição relativa das respostas. Na segunda questão a maior concentração das respostas está em “Nunca” com 41%, seguido por “Às vezes” com 38% e menor concentração em “Sempre” com 3%. Observa-se que entre os respondentes 55% não se sentem incapazes de controlar as coisas importantes em suas vidas e 7% se sentem incapazes.

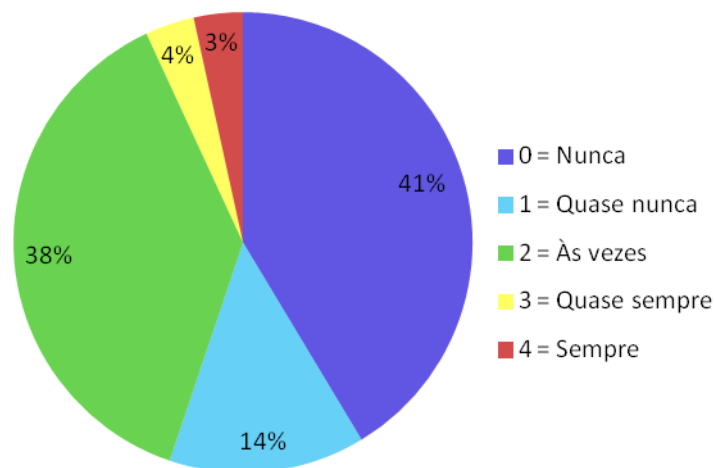


Figura 4.4 – Distribuição relativa (Questão 2)

A Figura 4.5 mostra os *clusters* gerados pela rede SOM para a terceira questão: “Você tem se sentido nervoso e “estressado?””.

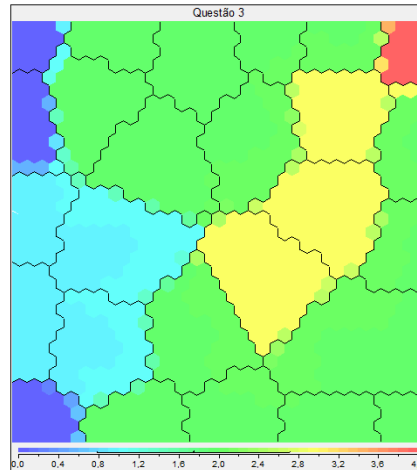


Figura 4.5 – *Clusters* (Questão 3)

A média aferida nessa questão foi de 1,76.

A Figura 4.6 mostra a distribuição relativa das respostas. Na terceira questão a maior concentração das respostas está em “Às vezes” com 59% e menor concentração em “Sempre” com 3%. Como se nota, 13% dos respondentes afirmam que se sentem nervosos e estressados. Isto confirma as afirmações de Coleman (1992), Primo (2008), Laranjeira (2009) e Umann (2011) sobre a perda de produtividade estar associada ao estresse.

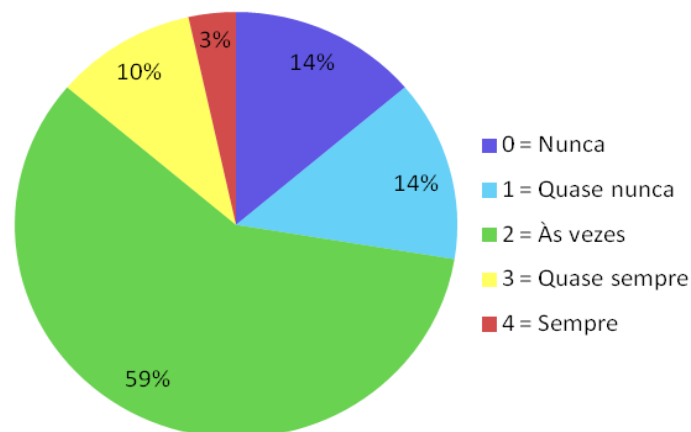


Figura 4.6 – Distribuição relativa (Questão 3)

A Figura 4.7 mostra os *clusters* gerados pela rede SOM para a quarta questão: “Você tem tratado com sucesso dos problemas difíceis da vida?”.

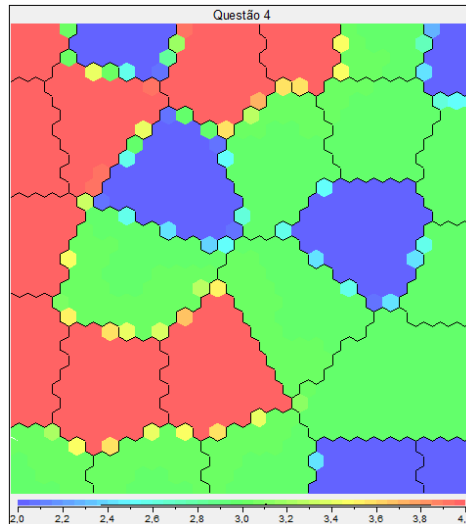


Figura 4.7 – *Clusters* (Questão 4)

A média aferida nessa questão foi de 3,07.

A Figura 4.8 mostra a distribuição relativa das respostas. Na quarta questão a maior concentração das respostas está em “Quase sempre” com 45% e nenhuma resposta em “Nunca” e “Quase nunca”. Desta Figura, conclui-se que 76% dos respondentes afirmam que sabem lidar com os problemas difíceis em suas vidas.

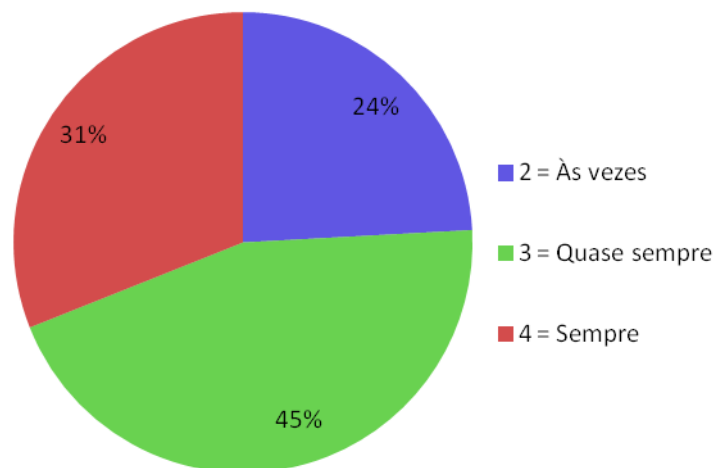


Figura 4.8 – Distribuição relativa (Questão 4)

A Figura 4.9 mostra os *clusters* gerados pela rede SOM para a quinta questão: “Você tem sentido que está lidando bem com as mudanças importantes que estão ocorrendo em sua vida?”.

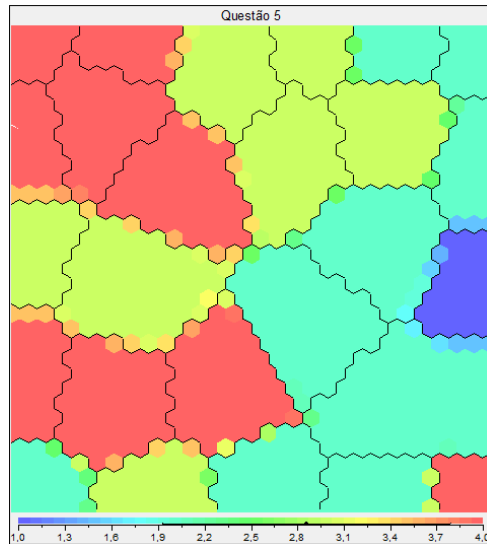


Figura 4.9 – *Clusters* (Questão 5)

A média aferida nessa questão foi de 2,83.

A Figura 4.10 mostra a distribuição relativa das respostas. Na quinta questão a maior concentração das respostas está em “Às vezes” com 41% e nenhuma resposta em “Nunca”. Como se percebe, 55% dos respondentes afirmam que sabem lidar com as mudanças em suas vidas e 4% não sabem.

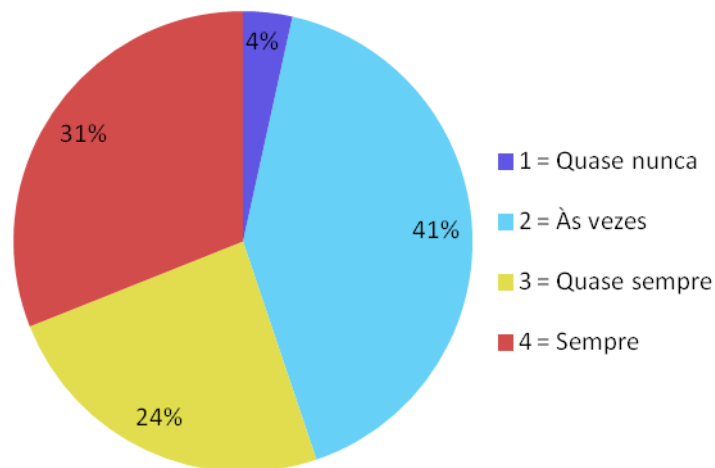


Figura 4.10 – Distribuição relativa (Questão 5)

A Figura 4.11 mostra os *clusters* gerados pela rede SOM para a sexta questão: “Você tem se sentido confiante na sua habilidade de resolver problemas pessoais?”.

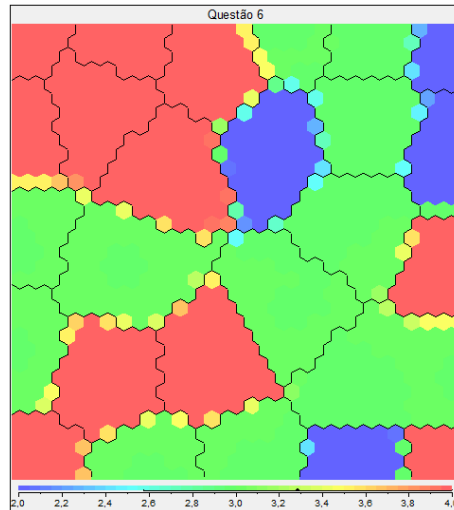


Figura 4.11 – *Clusters* (Questão 6)

A média aferida nessa questão foi de 3,24.

A Figura 4.12 mostra a distribuição relativa das respostas. Na sexta questão a maior concentração das respostas está em “Quase sempre” com 42%, seguida por 41% em “Sempre” e nenhuma resposta em “Nunca” e “Quase nunca”. Os resultados mostram que 83% dos respondentes sentem-se seguros para solucionar seus problemas pessoais.

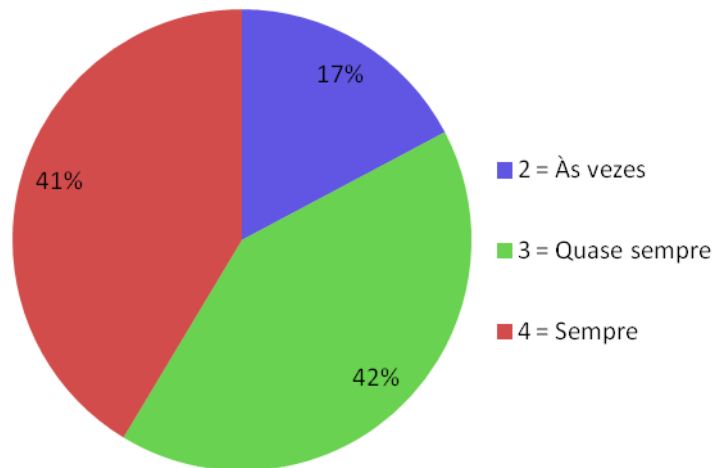


Figura 4.12 – Distribuição relativa (Questão 6)

A Figura 4.13 mostra os *clusters* gerados pela rede SOM para a sétima questão: “Você tem sentido que as coisas estão acontecendo de acordo com a sua vontade?”.

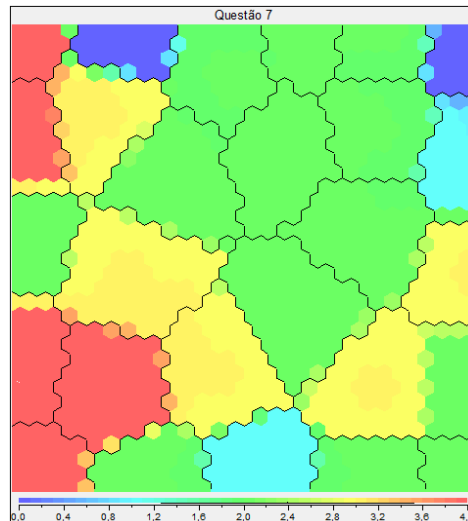


Figura 4.13 – *Clusters* (Questão 7)

A média aferida nessa questão foi de 2,38.

A Figura 4.14 mostra a distribuição relativa das respostas. Na sétima questão a maior concentração das respostas está em “Às vezes” com 48% e menor concentração em “Nunca” e “Quase nunca” ambas com 7%. Desta figura percebe-se que 14% dos respondentes não acreditam que as coisas estão acontecendo de acordo com suas vontades e 38% acreditam que sim.

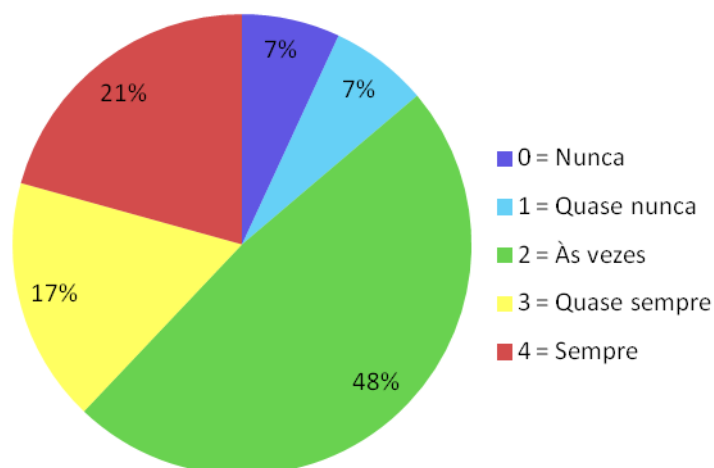


Figura 4.14 – Distribuição relativa (Questão 7)

A Figura 4.15 mostra os *clusters* gerados pela rede SOM para a oitava questão: “Você tem achado que não conseguiria lidar com todas as coisas que você tem que fazer?”.

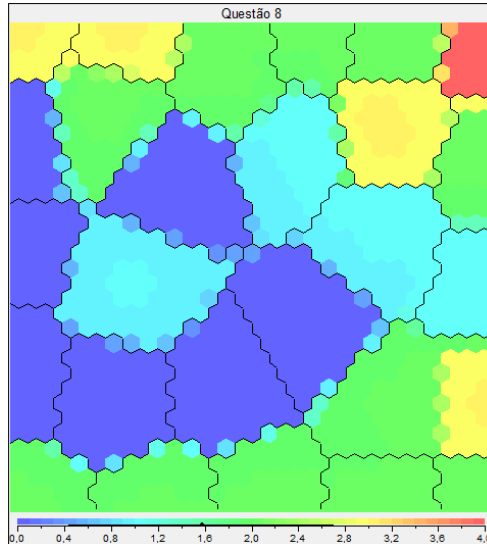


Figura 4.15 – *Clusters* (Questão 8)

A média aferida nessa questão foi de 1,59.

A Figura 4.16 mostra a distribuição relativa das respostas. Na oitava questão a maior concentração das respostas está em “Às vezes” com 45% e menor concentração em “Sempre” com 3%. Observa-se que 38% dos respondentes têm conseguido lidar com as suas tarefas e 17% não conseguem.

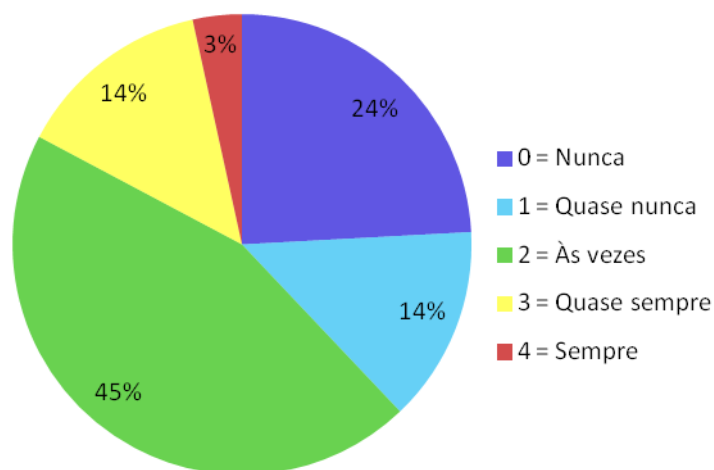


Figura 4.16 – Distribuição relativa (Questão 8)

A Figura 4.17 mostra os *clusters* gerados pela rede SOM para a nona questão: “Você tem conseguido controlar as irritações em sua vida?”.

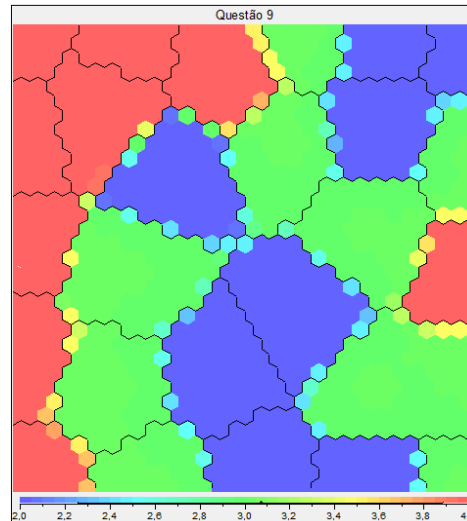


Figura 4.17 – *Clusters* (Questão 9)

A média aferida nessa questão foi de 3,03.

A Figura 4.18 mostra a distribuição relativa das respostas. Na nona questão as respostas estão bem distribuídas em “Às vezes”, “Quase sempre” e “Sempre” e nenhuma resposta em “Nunca” e “Quase nunca”. Nota-se que 69% dos respondentes conseguem controlar suas irritações.

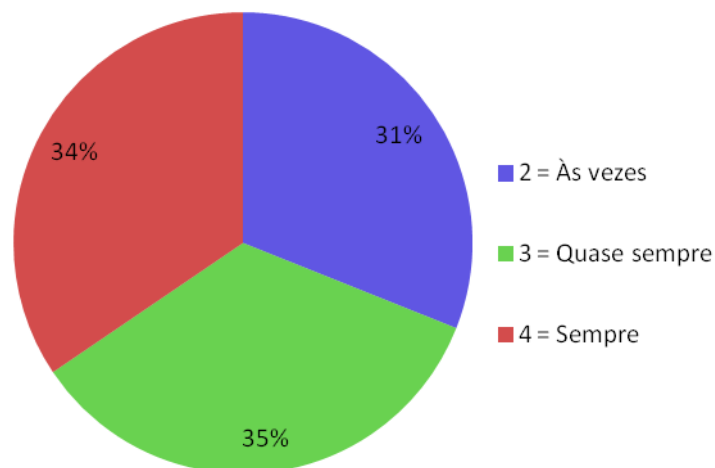


Figura 4.18 – Distribuição relativa (Questão 9)

A Figura 4.19 mostra os *clusters* gerados pela rede SOM para a décima questão: “Você tem sentido que as coisas estão sob o seu controle?”.

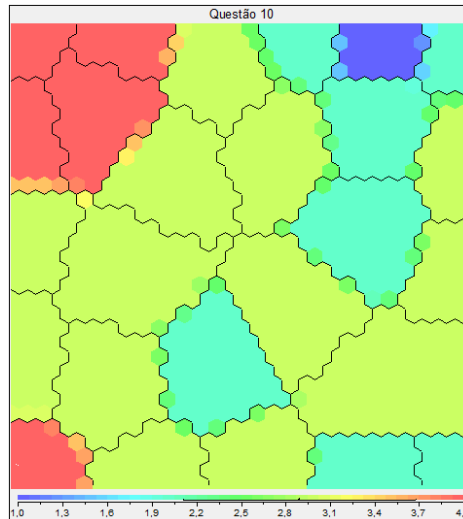


Figura 4.19 – *Clusters* (Questão 10)

A média aferida nessa questão foi de 2,86.

A Figura 4.20 mostra a distribuição relativa das respostas. Na décima questão a maior concentração das respostas está em “Quase sempre” com 48% e menor concentração em “Quase nunca)” com 3%, não houve resposta em “Nunca”. As respostas positivas corresponderam a 69% dos respondentes, o que está de acordo com as respostas para a segunda questão.

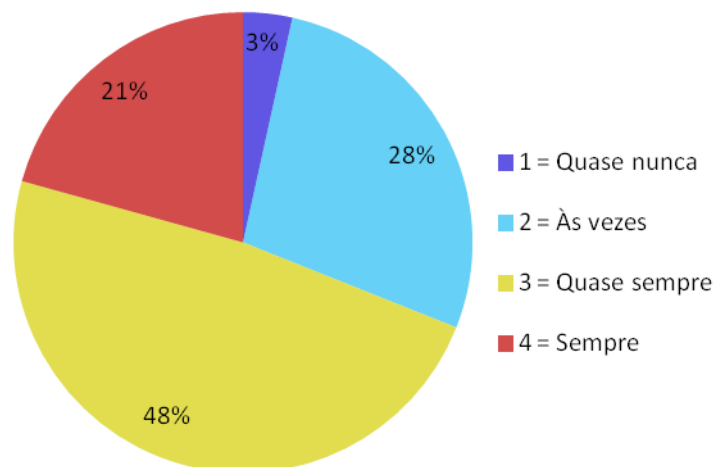


Figura 4.20 – Distribuição relativa (Questão 10)

A Figura 4.21 mostra os *clusters* gerados pela rede SOM para a décima primeira questão: “Você tem ficado irritado porque as coisas que acontecem estão fora do seu controle?”.

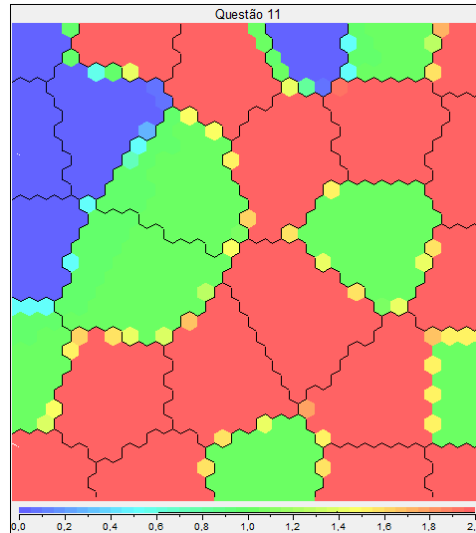


Figura 4.21 – *Clusters* (Questão 11)

A média aferida nessa questão foi de 1,41.

A Figura 4.22 mostra a distribuição relativa das respostas. Na décima primeira questão a maior concentração das respostas está em “Às vezes” com 59% e menor concentração em “Nunca” com 17%, não houve resposta em “Quase sempre” e “Sempre”. Como se nota 59% dos respondentes afirmam que às vezes ficam irritados com a falta de controle e 41% nunca ou quase nunca ficam irritados.

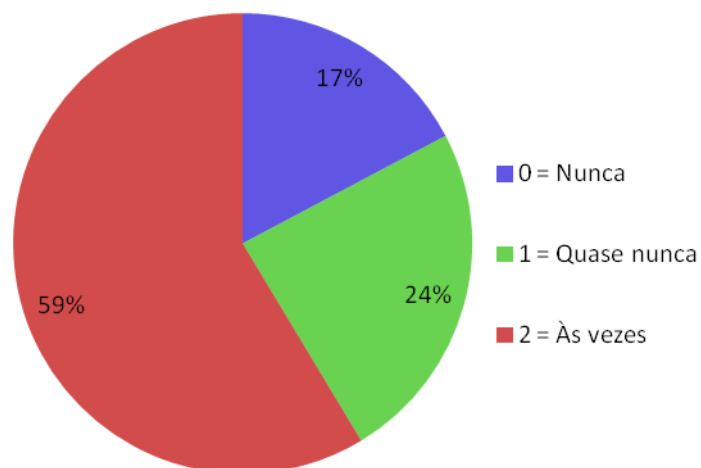


Figura 4.22 – Distribuição relativa (Questão 11)

A Figura 4.23 mostra os *clusters* gerados pela rede SOM para décima segunda questão: “Você tem se encontrado pensando sobre as coisas que deve fazer?”.

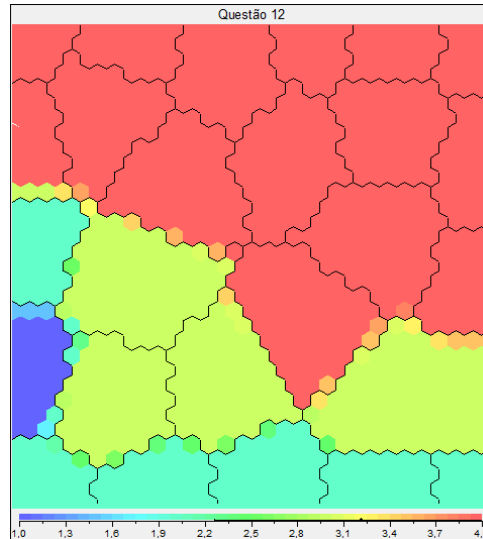


Figura 4.23 – *Clusters* (Questão 12)

A média aferida nessa questão foi de 3,17.

A Figura 4.24 mostra a distribuição relativa das respostas. Na décima segunda questão a maior concentração das respostas está em “Sempre” com 52% e menor concentração em “Quase nunca” com 3%, não houve resposta em “Nunca”. Da figura, 69% afirmam pensar sobre as coisas que deve fazer.

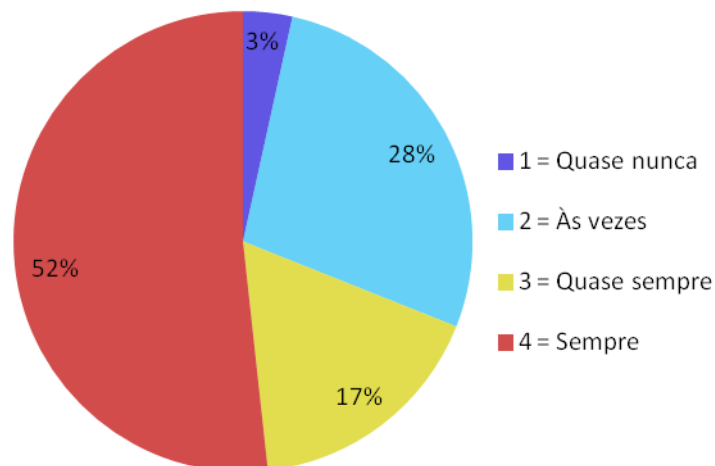


Figura 4.24 – Distribuição relativa (Questão 12)

A Figura 4.25 mostra os *clusters* gerados pela rede SOM para a décima terceira questão: “Você tem conseguido controlar a maneira como gasta seu tempo?”.

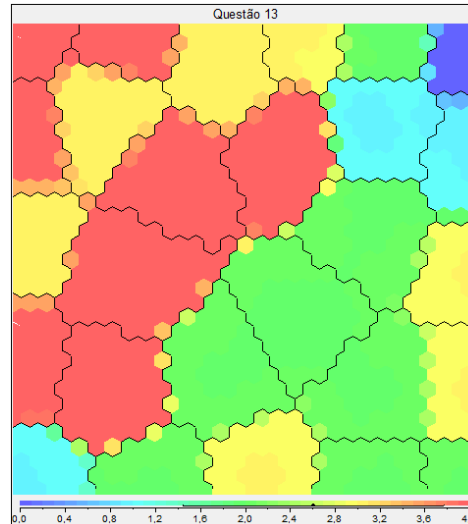


Figura 4.25 – *Clusters* (Questão 13)

A média aferida nessa questão foi de 2,59.

A Figura 4.26 mostra a distribuição relativa das respostas. Na décima terceira questão a maior concentração das respostas está em “Às vezes” com 31% e menor concentração em “Nunca” com 3%, seguida por “Quase nunca” com 14%. Desta questão pode-se notar que 52% dos respondentes acreditam que sabem controlar a maneira como gasta o seu tempo e 17% não conseguem controlar.

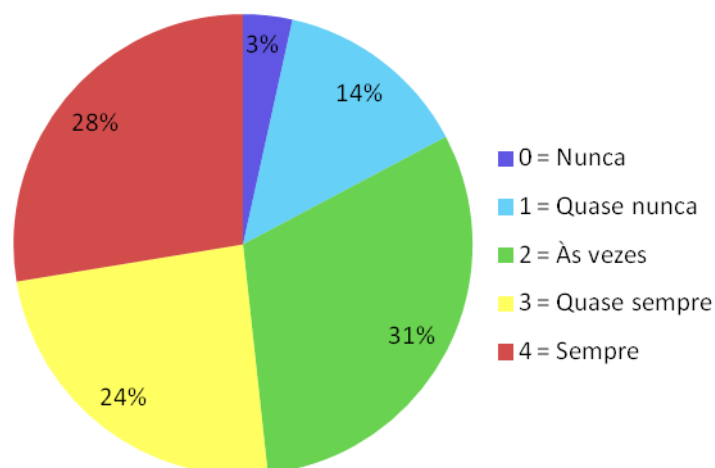


Figura 4.26 – Distribuição relativa (Questão 13)

A Figura 4.27 mostra os *clusters* gerados pela rede SOM para a décima quarta questão: “Você tem sentido que as dificuldades se acumulam a ponto de você acreditar que não pode superá-las?”.

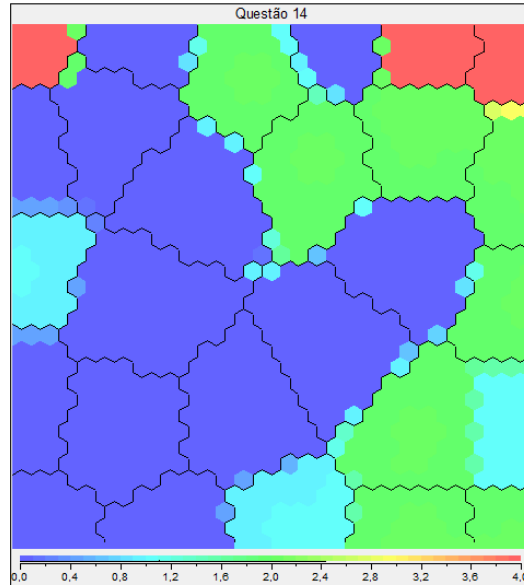


Figura 4.27 – Clusters (Questão 14)

A média aferida nessa questão foi de 1,14.

A Figura 4.28 mostra a distribuição relativa das respostas. Na décima quarta questão a maior concentração das respostas está em “Nunca” com 48% e menor concentração em “Sempre” com 10%, não houve resposta em “Quase sempre”. Observa-se que 59% dos respondentes afirmam que as dificuldades não estão se acumulando e podem superá-las e 41% não conseguem.

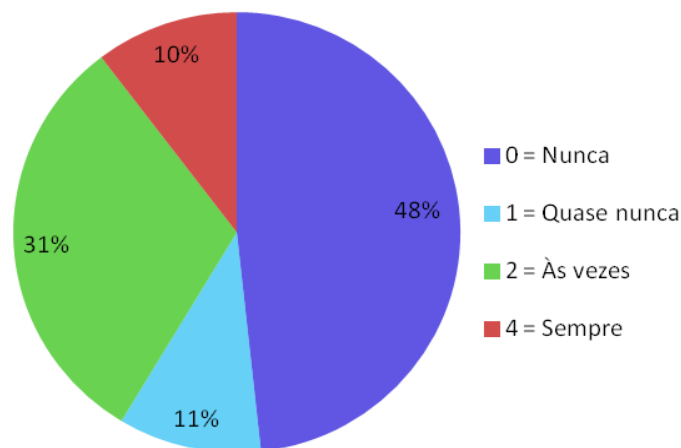


Figura 4.28 – Distribuição relativa (Questão 14)

Observa-se com os resultados obtidos através da *survey* que uma parcela dos empregados pesquisados, entre 4% a 11%, apresenta estresse o que indica um possível foco de presenteísmo, que em longo prazo pode causar absenteísmo prolongado.

4.2 REALIZAÇÃO DOS EXPERIMENTOS COM A REDE SOM

A base de dados, contendo as informações (todos os atributos) dos 36 empregados, é submetida à rede SOM para minerar os dados. A Figura 4.29 ilustra o mapa rotulado gerado pela rede SOM rotulada da seguinte forma:

- N (empregado com perfil dentro dos padrões Normais = 0)
- A (empregado com perfil Absenteísta = -1)
- P (empregado com perfil Presenteísta = 1)

A análise dos *clusters* de dados permite transmitir as informações através de um mapa visual intuitivo, que relaciona itens por similaridade, *clusters*, lacunas e *outliers* e podem ser ilustradas naturalmente usando relações espaciais, sombras e cores (KOHONEN, 2000).

Observa-se, Figura 4.29, que os empregados com perfil absenteísta 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.

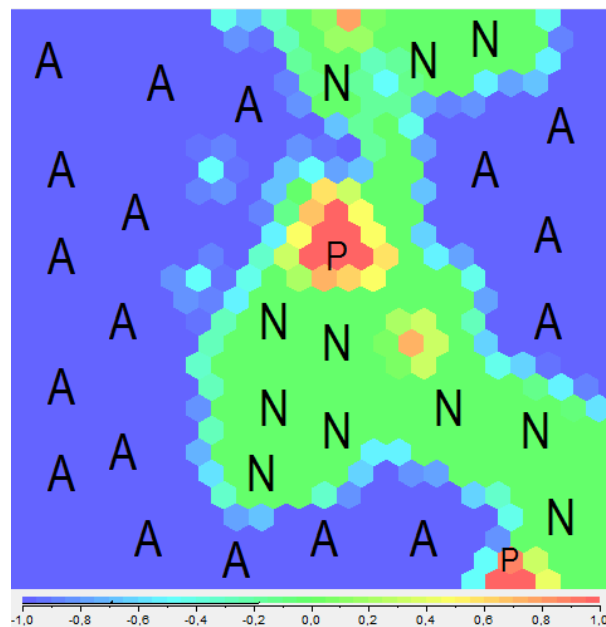


Figura 4.29 – Rede SOM rotulada

A Tabela 4.1 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.

Tabela 4.1 – Absenteísmo em horas

Rótulo	Absenteísmo em horas	(%)	Empregados por classe	(%)	Média de Absenteísmo em horas por empregado
A	3.638	71,00%	12	33,33%	303
N	1.193	23,28%	21	58,33%	57
P	293	5,72%	3	8,33%	98
Total	5.124	100,0%	36		

A realização dos experimentos segue agora evidenciando os atributos da base de dados dos empregados da empresa de *Courier*. As Figuras 4.30 até 4.39 ilustram os mapas gerados pela rede SOM.

A Figura 4.30 mostra os *clusters* gerados pela rede SOM (Atributo Mês).

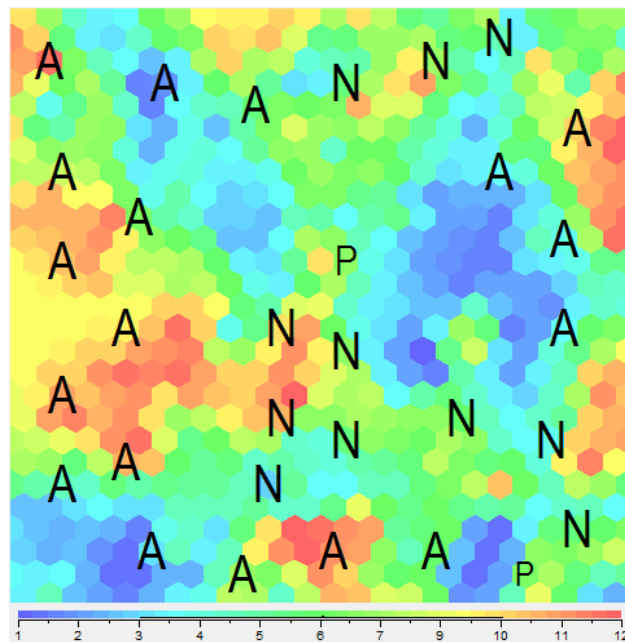


Figura 4.30 – Atributo Mês

Observa-se na Figura 4.30 e na Tabela 4.2 com o atributo mês (mês em que existe maior ou menor número de atestados), que existe maior absenteísmo nos meses de Março e Julho, que juntos correspondem a 29,2%, isso poderia indicar alguma doença sazonal ou algum tipo de atividade sazonal com maior risco de acidentes ou similares.

No mês de Janeiro constata-se o menor absenteísmo com 4,3%, seguido dos meses de Agosto, Fevereiro e Setembro com a média de 5,6%.

Tais comportamentos podem ser investigados com vistas à identificação do absenteísmo e prevenção. A Tabela 4.2 mostra a frequência simples e relativa relacionada ao absenteísmo em horas durante os meses do ano. Os meses de Março e Julho, que juntos representam 29,2% do absenteísmo e por esse motivo foram examinados separadamente,

demonstram a seguinte relação: empregados absenteístas (A) representam 66% do total do absenteísmo em horas, empregados dentro dos padrões normais (N) representam 17% e os empregados presenteístas também representam 17%, causando surpresa, pois o absenteísmo desse grupo de empregados nesses dois meses corresponde a 86%.

Tabela 4.2 – Meses *versus* Absenteísmo

Mês	Absenteísmo em horas	(%)	(%) Acumulada
Janeiro	222	4,3%	4,3%
Fevereiro	294	5,7%	10,1%
Março	765	14,9%	25,0%
Abril	482	9,4%	34,4%
Maio	400	7,8%	42,2%
Junho	411	8,0%	50,2%
Julho	734	14,3%	64,6%
Agosto	288	5,6%	70,2%
Setembro	292	5,7%	75,9%
Outubro	349	6,8%	82,7%
Novembro	473	9,2%	91,9%
Dezembro	414	8,1%	100,0%
Total	5124		

Observa-se na Figura 4.31 e na Tabela 4.3 com o atributo dia da semana, que o absenteísmo concentra-se nas segundas e terças-feiras, juntas correspondem a 53% do absenteísmo, ou seja, no início da semana, esse comportamento pode ser investigado, pois pode representar a falta de interesse ou desmotivação com as tarefas diárias. Outra preocupação pode ser atribuída aos excessos no final de semana, por exemplo. Constatase que o menor absenteísmo ocorre na quinta-feira com 11%, seguida da sexta-feira com 14%.

A Figura 4.31 mostra os *clusters* gerados pela rede SOM (Atributo Dia da Semana).

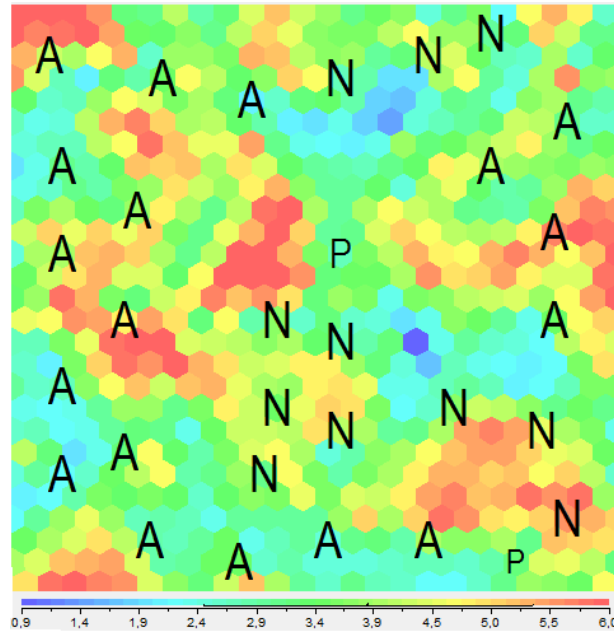


Figura 4.31 – Atributo Dia da Semana

A Tabela 4.3 mostra a frequência simples e relativa relacionada ao absenteísmo em horas durante os dias semana.

Tabela 4.3 – Dia da Semana *versus* Absenteísmo

Dia da Semana	Absenteísmo em horas	(%)	(%) Acumulada
Segunda	1.489	29%	29%
Terça	1.229	24%	53%
Quarta	1.115	22%	75%
Quinta	553	11%	86%
Sexta	738	14%	100%
Total	5.124	100%	

Observa-se na Figura 4.32 e na Tabela 4.4 com o atributo Estação do ano, que o absenteísmo tem maior concentração no Outono com 29,1%, e pode indicar doenças ligadas com o tipo de atividade e o clima. Esse conhecimento pode subsidiar a tomada de decisão no sentido de viabilizar programas de controle e prevenção de doenças por parte da empresa, em determinadas estações do ano. O menor absenteísmo ocorre no Verão com 22,5%.

Legenda: 1 = Verão, 2 = Outono, 3 = Inverno, 4 = Primavera, 0 = Empregados sem nenhum registro de absenteísmo durante a pesquisa, o que pode ser interpretado inicialmente como comportamento Presenteísta.

A Figura 3.33 mostra os *clusters* gerados pela rede SOM (Atributo Estação do Ano).

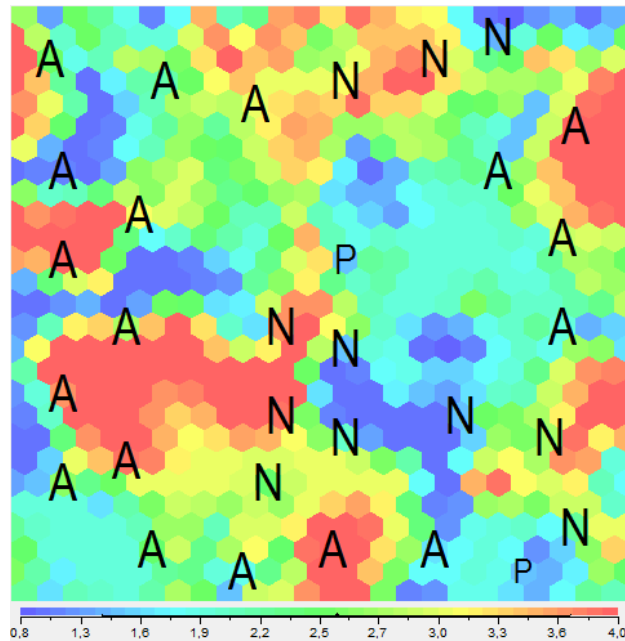


Figura 4.32 – Atributo Estação do Ano

A Tabela 4.4 mostra a frequência simples, relativa e acumulada relacionada ao absenteísmo em horas durante as Estações do Ano.

Tabela 4.4 – Estação do Ano *versus* Absenteísmo

Estações do Ano	Absenteísmo em horas	(%)	(%) Acumulada
Verão	1.152	22,5%	22,5%
Outono	1.492	29,1%	51,6%
Inverno	1.241	24,2%	75,8%
Primavera	1.239	24,2%	100,0%
Total	5.124		

Observa-se na Figura 4.33 e na Tabela 4.5 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. Consta-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 4.33 mostra os *clusters* gerados pela rede SOM (Atributo Tempo de Afastamento em horas).

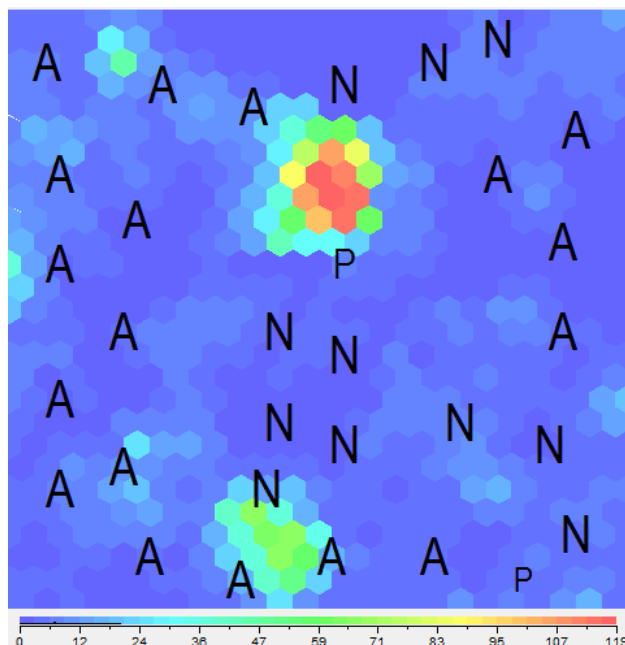


Figura 4.33 – Atributo Tempo de Afastamento em Horas

A Tabela 4.5 mostra a frequência simples e relativa relacionada aos registros e ao absenteísmo em horas.

Tabela 4.5 – Tempo de Afastamento em Horas

Classe	Registros	(%)	Absenteísmo em horas	(%)
Até 20 horas	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	100,00%	5.124	100,0%

Verifica-se na Figura 4.34 e na Tabela 4.6 com o atributo Distância da Residência para o Trabalho em Quilômetros, que a maior parte dos absenteístas, reside 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 3.35 mostra os *clusters* gerados pela rede SOM (Atributo Distância da Residência para o Trabalho em quilômetros).

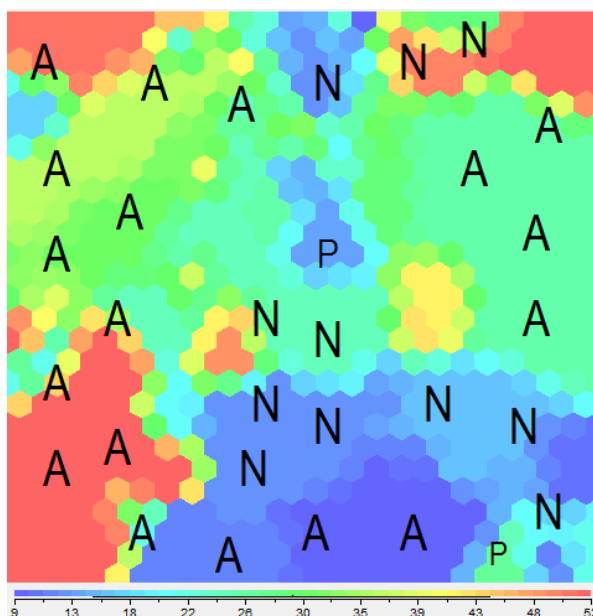


Figura 4.34 – Atributo Distância da Residência para o Trabalho em Quilômetros

A Tabela 4.6 mostra a frequência simples e relativa relacionada à distância da residência para o trabalho em quilômetros, empregados, registros e o absenteísmo em horas.

Tabela 4.6 – Distância da Residência para o Trabalho em Quilômetros

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
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	100,00%	5124	100,0%

Verifica-se na Figura 4.35 e na Tabela 4.7 com o atributo Tempo de Serviço, que os empregados com maior tempo de serviço apresentam comportamento Normal (N) e Presenteísta (P), e que os empregados com menor tempo de serviço apresentam comportamento Normal.

Constata-se que a maior parte dos absenteístas tem entre 16 e 20 anos de serviço, esse grupo corresponde a 13,9% dos empregados, contudo, representam 28,0% do absenteísmo em horas.

A Figura 4.35 mostra os *clusters* gerados pela rede SOM (Atributo Tempo de serviço).

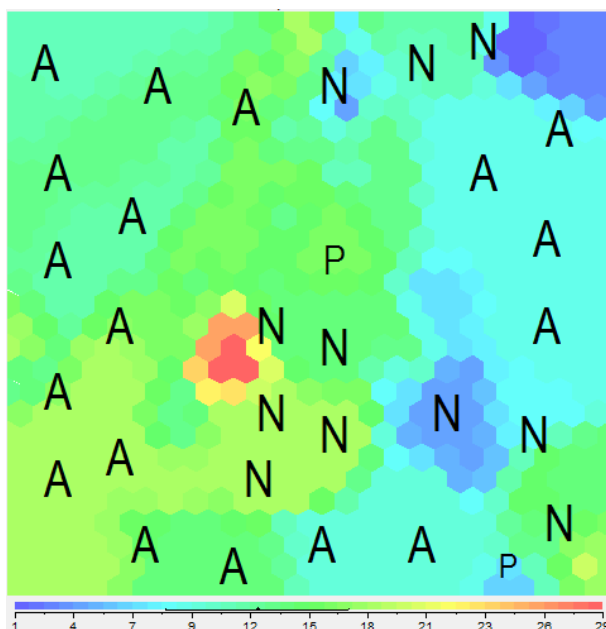


Figura 4.35 – Atributo Tempo de Serviço

A Tabela 4.7 mostra a frequência simples e relativa relacionada ao tempo de serviço, empregados, registros e o absenteísmo em horas.

Tabela 4.7 – Tempo de Serviço

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
Até 5 anos	3	8,3%	47	6,4%	338	6,6%
De 6 a 10 anos	8	22,2%	208	28,1%	1076	21,0%
De 11 a 15 anos	18	50,0%	273	36,9%	2243	43,8%
De 16 a 20 anos	5	13,9%	205	27,7%	1435	28,0%
De 21 anos em diante	2	5,6%	7	0,9%	32	0,6%
Total	36	100,0%	740	100,00%	5124	100,0%

Constata-se na Figura 4.36 e na Tabela 4.8 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 4.36 mostra os *clusters* gerados pela rede SOM (Atributo Idade).

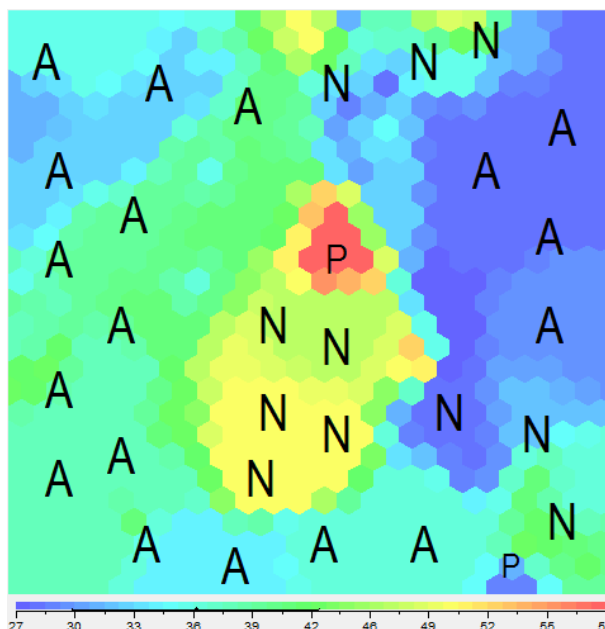


Figura 4.36 – Atributo Idade

A Tabela 4.8 mostra a frequência simples e relativa relacionada à Idade, empregados, registros e o absenteísmo em horas.

Tabela 4.8 – Idade *versus* Absenteísmo

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
Até 30 anos	6	16,7%	177	23,9%	962	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	100,00%	5124	100,0%

Constata-se na Figura 4.37 e na Tabela 4.9 com o atributo Escolaridade, que a 85,7% do absenteísmo em horas pertence à classe com ensino médio. Verifica-se ainda que à medida que a escolaridade aumenta a relação entre o empregado e absenteísmo em horas diminui.

Legenda: 1 = Ensino médio, 2 = Cursando Ensino Superior, 3 = Ensino Superior completo, 4 = Pós-graduado / graduando.

A Figura 4.37 mostra os *clusters* gerados pela rede SOM (Atributo Escolaridade).

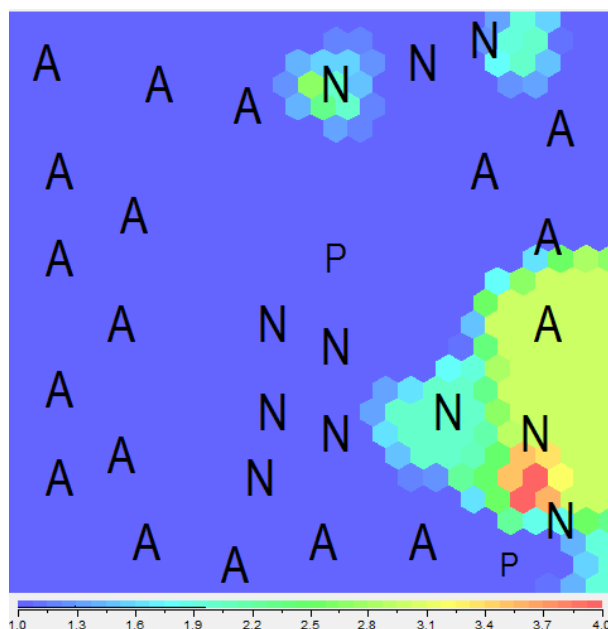


Figura 4.37 – Atributo Escolaridade

A Tabela 4.9 mostra a frequência simples e relativa relacionada à Escolaridade, empregados, registros e o absenteísmo em horas.

Tabela 4.9 – Escolaridade *versus* Absenteísmo

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
Ensino Médio	28	77,8%	610	82,4%	4393	85,7%
Graduando	4	11,1%	46	6,2%	294	5,7%
Graduado	3	8,3%	80	10,8%	416	8,1%
Pós-graduando/graduado	1	2,8%	4	0,5%	21	0,4%
Total	36	100,0%	740	100,00%	5124	100,0%

Na Figura 4.38 e na Tabela 4.10 verifica-se com o atributo número de filhos, que o comportamento absenteísta é muito parecido entre os empregados que não tem filhos, e os que têm até dois filhos. O comportamento do empregado com 4 filhos mostra uma relação maior com o absenteísmo.

A Figura 4.38 mostra os *clusters* gerados pela rede SOM (Atributo Número de filhos).

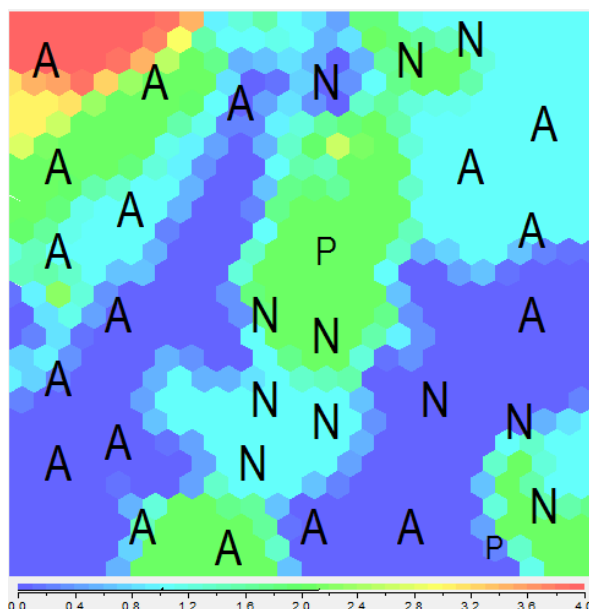


Figura 4.38 – Atributo Número de filhos

A Tabela 4.10 mostra a frequência simples e relativa relacionada ao número de filhos, empregados, registros e o absenteísmo em horas.

Tabela 4.10 – Número de filhos x Absenteísmo

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
Número de Filhos (0)	12	33,3%	298	40,3%	1605	31,3%
Número de Filhos (1)	10	27,8%	229	30,9%	1381	27,0%
Número de Filhos (2)	12	33,3%	156	21,1%	1649	32,2%
Número de Filhos (3)	1	2,8%	15	2,0%	183	3,6%
Número de Filhos (4)	1	2,8%	42	5,7%	306	6,0%
Total	36	100,0%	740	100,00%	5124	100,0%

Na Figura 4.39 e na Tabela 4.11 verifica-se com o atributo Índice de Massa Corporal, que o comportamento absenteísta encontra-se nos empregados com IMC entre 30,0 a 34,9 (Obesidade grau I), já entre os empregados com IMC entre 18,6 a 24,9 apresentam o melhor resultado com o absenteísmo.

A Figura 4.39 mostra os *clusters* gerados pela rede SOM (Atributo Índice de Massa Corporal).

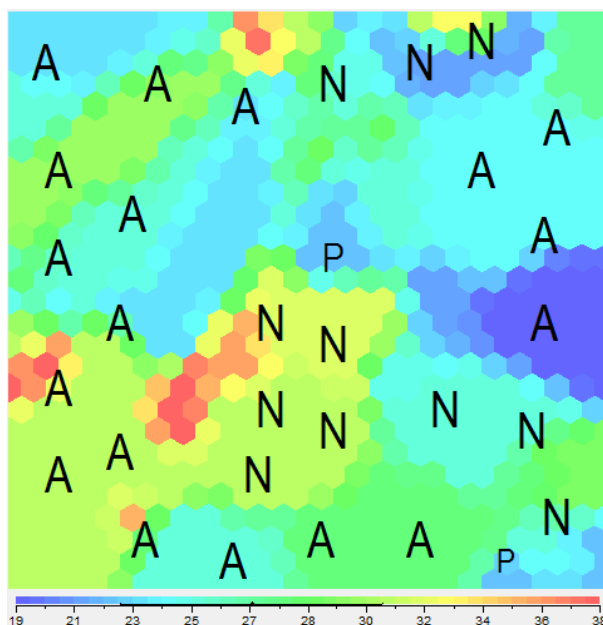


Figura 4.39 – Atributo Índice de Massa Corporal

A Tabela 4.11 mostra a frequência simples e relativa relacionada ao Índice de Massa Corporal, empregados, registros e o absenteísmo em horas.

Tabela 4.11 – Índice de Massa Corporal *versus* Absenteísmo

Classe	Empregados	(%)	Registros	(%)	Absenteísmo em horas	(%)
Saudável (18,6 – 24,9)	19	52,8%	331	44,7%	2537	49,5%
Sobrepeso (25,0 – 29,9)	8	22,2%	165	22,3%	1126	22,0%
Obesidade Grau I (30,0 – 34,9)	7	19,4%	220	29,7%	1341	26,2%
Obesidade Grau II (35,0 – 39,9)	2	5,6%	24	3,2%	120	2,3%
Total	36	100,0%	740	100,00%	5124	100,0%

4.2.1 REALIZAÇÃO DOS EXPERIMENTOS COM A REDE MLP

Pode-se verificar nos mapas gerados pela rede SOM que os *clusters* são de boa qualidade, porque apresentam bordas definidas, erro de quantização dentro dos limites (0,7712) e erro topográfico baixo (0,087).

A confiabilidade dos Clusters pode ser verificada ao classificar, em uma rede MLP, os empregados rotulados no mapa.

Observa-se, que na fase de teste, Figura 4.40, que a rede MLP classificou corretamente 93% dos empregados absenteístas e presenteístas. Essa taxa de classificação demonstra que os *clusters* gerados pela rede SOM são confiáveis.

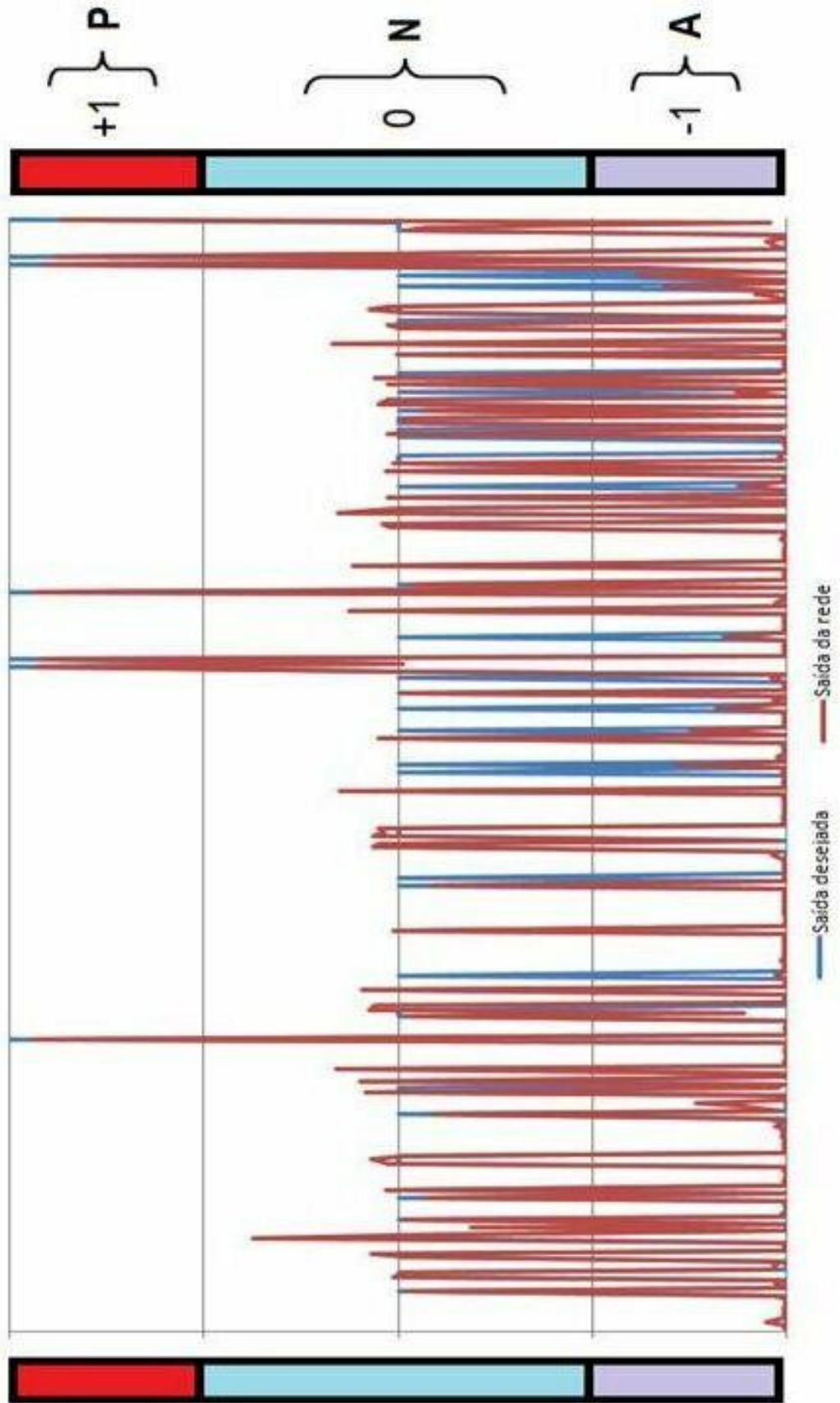


Figura 4.40 – Resultado da MLP na classificação da rede SOM rotulada

Na classificação realizada pela rede MLP foi considerado como classificação correta os erros quadrados menores que 0,5.

4.3 DISCUSSÃO DOS RESULTADOS

Apresenta-se, 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 absenteísta, presenteísta e dentro do padrão normal contido na base de dados da empresa de *courier* estudada. A Tabela 4.12 apresenta de forma sumarizada os resultados.

Tabela 4.12 – Sumário dos perfis estudados

Atributos	Absenteísta (A)	Normal (N)	Presenteísta (P)
Absenteísmo em horas (%)	71%	23,33%	5,70%
Empregados (%)	33,30%	58,33%	8,33%
Mês (maior ocorrência de empregados)	Julho / Março	Abril / Junho	Julho / Março
Mês (menor ocorrência de empregados)	Agosto	Setembro	Jan/Abr/Jun/Set/Nov
Dia da semana (maior ocorrência de empregados)	Segunda	Quarta	Terça
(menor ocorrência de empregados)	Quinta	Terça	Quinta
Estação do ano (maior ocorrência de empregados)	Primavera	Outono	Inverno
Estação do ano (menor ocorrência de empregados)	Verão	Primavera	Primavera
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
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
Tempo de serviço (maior ocorrência de empregados)	11 a 15 anos	11 a 15 anos	16 a 20 anos
Tempo de serviço (menor ocorrência de empregados)	1 a 5 anos	21 a 29 anos	6 a 10 anos
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
Escolaridade	93% ensino médio 7% graduados	60% ensino médio 25% graduando 2% graduado/pós-graduado	100% ensino médio
Nº de filhos (maior ocorrência de empregados)	0	1	2
Nº de filhos (menor ocorrência de empregados)	3	0	0
Índice de massa corporal (IMC)	Entre 19,2 -24,9 e 30,0 - 34,9	Entre 19,2 -24,9 e 30,0 - 34,9	Entre 19,2 -24,9

Apresenta-se nos itens a, b e c as características dos perfis dos empregados estudados.

a) Características do Perfil Absenteísta:

- este perfil corresponde a 71% das horas perdidas e a 33,33% dos empregados com maior incidência do absenteísmo nos meses de Julho e Março e menor incidência no mês de Agosto, as horas perdidas se concentram às segundas-feiras, a estação do ano de maior incidência é a Primavera e a menor incidência o Verão;

- com relação ao tempo de afastamento, o perfil apresenta maior incidência de absenteísmo em até 20 horas e menor incidência em até 60 horas;

- em se tratando da distância da residência para o trabalho em quilômetros, o perfil apresenta maior incidência acima de 41 quilômetros de distância do trabalho e menor incidência até 10 quilômetros de distância do trabalho;

- com relação ao tempo de serviço, o perfil apresenta maior incidência de absenteísmo na faixa dos 11 aos 15 anos e com menor incidência na faixa de 1 a 5 anos;

- com relação à idade dos empregados, o perfil 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;

- quando se trata de escolaridade, esse perfil é composto basicamente por empregados com ensino médio completo igual a 93% e graduados com 7% dos empregados;

- quanto ao número de filhos, o perfil apresenta maior incidência de absenteísmo entre os empregados que não possuem filhos e menor incidência entre os que possuem 3 filhos; e

- no caso de índice de massa corporal, o perfil apresenta maior incidência entre os empregados considerados saudáveis (IMC – 19,2 – 24,9), seguidos por empregados com obesidade grau I (IMC - 30,0 – 34,9).

b) Características do Perfil Presenteísta:

- este perfil corresponde a 5,7% das horas perdidas e 8,33% dos empregados com maior incidência de absenteísmo nos meses de Julho e Março e nenhuma incidência nos meses de Janeiro, Abril, Junho, Setembro e Novembro, as horas perdidas se concentram às terças-feiras, a estação do ano de maior incidência é o Inverno e a menor incidência a Primavera;

- com relação ao tempo de afastamento, o perfil apresenta maior incidência de absenteísmo em até 120 horas e menor incidência em até 20 horas;

- em se tratando da distância da residência para o trabalho em quilômetros, o perfil apresenta maior incidência em até 20 quilômetros de distância do trabalho e com menor incidência em até 30 quilômetros de distância do trabalho;

- quanto ao tempo de serviço, o perfil apresenta maior incidência de Absenteísmo na faixa dos 16 aos 20 anos e com menor incidência na faixa de 6 a 10 anos;
- com relação à idade dos empregados, o perfil apresenta maior incidência de absenteísmo na faixa de 46 a 58 anos de idade e menor incidência até 30 anos de idade;
- em se tratando de escolaridade, o perfil compõe-se apenas por empregados com ensino médio completo;
- quanto ao número de filhos, o perfil apresenta maior incidência de absenteísmo entre os empregados que possuem 2 filhos e menor incidência entre os que não possuem filhos; e
- no caso de índice de massa corporal, o perfil apresenta apenas incidência entre os empregados considerados saudáveis (IMC – 19,2 – 24,9).

Verifica-se também que a maior incidência do absenteísmo ocorre nos meses de Julho e Março, justamente quando os empregados presenteístas também têm maior incidência de absenteísmo. Isso pode ter uma relação com a sobrecarga de trabalho que é distribuída entre os empregados que estão trabalhando e apresentam o perfil presenteísta.

c) Características do Perfil dentro do padrão normal:

- este perfil corresponde a 23,3% das horas perdidas e 58,33% dos empregados com maior incidência de absenteísmo nos meses de Abril e Junho e menor incidência no mês de Setembro, as horas perdidas se concentram às quartas-feiras, a estação do ano de maior incidência é o Outono e a de menor incidência a Primavera;
- com relação ao tempo de afastamento, esse perfil apresenta maior incidência de absenteísmo em até 20 horas e menor incidência em até 40 horas;
- em se tratando da distância da residência para o trabalho em quilômetros, o perfil apresenta maior incidência em até 20 quilômetros de distância do trabalho e com menor incidência em até 10 quilômetros de distância do trabalho;
- com relação ao tempo de serviço, o perfil apresenta maior incidência de absenteísmo na faixa dos 11 aos 15 anos e com menor incidência na faixa de 21 a 29 anos;
- com relação à idade dos empregados, esse perfil 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 se tratando de escolaridade, o perfil é composto por empregados com ensino médio completo igual a 60%, graduandos com 25%, graduados com 14% e pós-graduando e pós-graduado com 2% dos empregados;

- quanto ao número de filhos, esse perfil apresenta maior incidência de absenteísmo entre os empregados que possuem 1 filho e menor incidência entre os que não possuem filhos;
e

- no caso de índice de massa corporal, esse perfil apresenta maior incidência entre os empregados considerados saudáveis (IMC – 19,2 – 24,9), seguidos por empregados com obesidade grau I (IMC - 30,0 – 34,9).

5 CONCLUSÃO

Este estudo identificou o perfil dos empregados absenteístas e presenteístas em uma empresa de *Courier* da cidade de São Paulo, aplicando como técnica de mineração de dados uma rede neural artificial do tipo Mapa Auto-Organizável de Kohonen ou *Self-Organizing Map* (rede SOM).

Os agrupamentos gerados pela rede SOM apresentaram boa qualidade em função do erro de quantização e topográfico baixo, da boa visualização e das bordas bem definidas. Em seguida, os registros agrupados pela rede SOM foram submetidos à classificação através de uma rede neural artificial do tipo *Multilayer Perceptron* (MLP), que classificou corretamente 93% dos registros, o que confirma a qualidade dos agrupamentos.

Obteve-se, com a mineração de dados, conhecimento útil sobre a base de dados dos funcionários da empresa de *Courier*, descritos a seguir:

- o perfil absenteísta representou cerca de trinta por cento dos empregados e o perfil presenteísta cerca de oito por cento, já o perfil dentro do padrão normal representou cerca de sessenta por cento dos empregados.

- quanto às horas perdidas de trabalho verificou-se que cerca de setenta por cento foi gerada pelos empregados absenteístas, os presenteístas foram responsáveis por cerca de seis por cento, enquanto os empregados dentro do padrão normal cerca de vinte por cento.

- quanto aos dias da semana, nas segundas-feiras a maior ocorrência pertenceu ao perfil absenteísta, já o perfil presenteísta apresentou maior ocorrência às terças-feiras, os empregados com perfil dentro do padrão normal apresentaram maior ocorrência às quartas-feiras.

- a distância do local de trabalho e da residência teve grande influência no perfil absenteísta. Percebeu-se também que o empregado com tempo de serviço maior e mais idoso tende ao presenteísmo, esse grupo também chamou a atenção pela baixa escolaridade.

O conhecimento útil extraído pela mineração de dados pode contribuir na tomada de decisão da empresa de *Courier* nos seguintes aspectos:

- redução dos 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;

- 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 propostas de continuidade do trabalho são:

- estudo da relação entre ergonomia, absenteísmo e presenteísmo;
- análise detalhada do custo do absenteísmo e do presenteísmo;
- estudo das características comportamentais que levam a um dos dois perfis; e
- busca de empregados que podem apresentar os dois perfis.

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 empresas de diversos setores.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFFONSO, C. Aplicação de Redes Neuro *Fuzzy* ao Processamento de Polímeros na Indústria Automotiva. Dissertação de Mestrado em Engenharia de Produção – Universidade Nove de Julho. 110 p., São Paulo, 2010.
- AGUIAR, G. A. S.; OLIVEIRA, J. R. Absenteísmo: Suas principais causas e conseqüências em uma empresa do ramo de saúde. *Revista de Ciências Gerenciais*, vol. 13, n.18, 2009.
- ALVES, M. A.; GODOY, S. C. B.; SANTANA, D. M. Motivos de licenças médicas em um hospital de urgência-emergência. *Revista Brasileira de Enfermagem*, Brasília, vol. 59, n. 2, p. 195-200, mar./abr. 2006.
- ANUMALLA, K. Sistema de Gestão da Pré-Processamento de Dados. Dissertação de Mestrado em Ciência – Universidade de Akron, USA, 64 p. Akron, 2007.
- AZEVEDO, C. S.; SANTOS, M. F. *Data Mining - Descoberta de Conhecimento em Bases de Dados*. Lisboa, Portugal: FCA – ed. Informática – Coleção: Sistema de Informação, 214 p., 2005.
- BACHOUCH, R. B.; GUINET, A.; HAJRI-GABOUJ, S. *An optimization model for task assignment in home health care. Health Care Management (WHCM), IEEE Workshop on., Venice, 2010.*
- BARROS, A. S.; CAMPOS, F. C. 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. 2006.
- BERGSTROM, G.; BODIN, L.; HAGBERG, J.; LINDH, T.; ARONSSON, G.; JOSEPHSON, M. *Does sickness presenteeism have an impact on future general health? Int Arch Occup Environ Health*, 2009.
- BERRY, M. J. A.; LINOFF, G. *Data Mining Techniques: For Marketing, Sales, and Customer Support*. New York: Wiley Computer Publishing, 1997.
- BIGUS, J. P. *Data Mining with Neural Network: Solving Business Problems from Applications Development to Decision Support*. Mcgraw-Hill, 1996.
- BONVENTTI, R. C. Projetos Saudáveis. Rev. Melhor. Disponível em: < <http://revistamelhor.uol.com.br/textos/255/artigo223482-1.asp>>. Acesso em: 10 jan. 2012.
- BLAZADONAKIS, E.; MICHALIS, Z. *Support Vector Machines and Neural Networks as Marker Selectors in Cancer Gene Analysis. In: Intelligent Techniques and Tools for Novel System Architectures*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, p. 237-258, 2008.

- BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro. Ed. LTC, 2ª ed. Rio de Janeiro, 2011.
- BUCKERMAN, P; LAUKKANEN, E. *Presenteeism in Finland: Determinants by Gender and the sector of conomy*. *Ege Akademik Bakış / Ege Academic Review* 9 (3): p.1007-1016, 2009.
- CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A. *Discovering Data Mining: from concept to implementation*. New Jersey : Prentice Hall, 1998.
- CALGAROTTO, R.; PINHEIRO, L. R. S. *Empresa, Colaborador e Ausência: um relato de experiência sobre o Absenteísmo nas relações de trabalho*. *Perspeciva*, Erechim. vol. 34, n.128, p. 167-175, dez. 2010.
- CANFIELD, G. W.; SOASH, D. G. *Presenteeism- A constructive view*. *Industrial Medicine and Surgery*, vol. 24, p. 417-418, 1955.
- CANTO, W. F.; SOUZA, L. S.; CANTO, N. C. F. *Data mining: An application to the semiconductor industry*. (CISTI), 2011 - *IEEE Proceedings of the 6th Conference Information Systems and Technologies Iberian*. p. 1-4, 2011.
- CARVALHO, L. A. V. *Redes Neurais Artificiais: a volta do cérebro eletrônico?* In: *Revista Ciência Hoje*. São Paulo, volume 12, n. 70, p.12-21, jan./fev. 1991.
- CARVALHO, L. A. V. *Data Mining: a Mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração*. Editora Ciência Moderna Ltda. Rio de Janeiro, 2005.
- CASCIO, W.; BOUDREAU, J. *Investing in People: Financial Impact of Human Resource Initiatives*, Pearson Education, Inc. Publishing as FT Press, New Jersey, 2008.
- CHIAVENATO, I. *Recursos humanos na empresa*. 3ª ed. São Paulo: Atlas, 1994.
- CHIAVENATO, I. *A administração de recursos humanos*. Recursos Humanos. 4ª ed. São Paulo: Atlas, 1997.
- CHIAVENATO, I. *Gestão de Pessoas: o novo papel dos recursos humanos nas organizações*. Rio de Janeiro: Elsevier, 1999.
- CHIAVENATO, I. *Gestão de Pessoas*. Edição compacta. 2ª ed. São Paulo: Atlas, 2004.
- COLEMAN, Vernon. *Técnicas de controle de estresse: como administrar a saúde das pessoas para aumentar os lucros*. Rio de Janeiro, 1992.
- COSTA, C. W. G. S.; COSTA, D. S. *O Stress Ocupacional como fator Interveniente na Produtividade Organizacional*. XXIII Encontro Nacional de Engenharia de Produção – ENEGEP. Ouro Preto, MG. p.1-8, 2003.
- COSTA, J. A. F.; ANDRADE NETTO, M. L. *Segmentação de Mapas Auto-Organizáveis com Espaço de Saída 3-D*. *Revista Controle & Automação*/Vol.18 n.2/abr, mai. e Jun. 2007.

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.

DANATRO, D. *Ausentismo Laboral de causa médica em uma Institución pública de Montevideo*: 1994-1995. *Revista Medica del Uruguay*, 13(2): p. 101-109, ago. 1997.

DALL'INHA, G. R. A Influência das Práticas e das Políticas de Recursos Humanos Sobre o Absenteísmo e a Rotatividade: um Estudo de Caso. Dissertação de Mestrado em Gestão da Qualidade e Produtividade – Universidade Federal de Santa Catarina. 92 p., Florianópolis/SC 2006.

DATASUS, Versão 1.6c em Português, 2000.

DIACOV, N.; LIMA, J. R. S. Absenteísmo Odontológico. *Revista Odontológica*. UNESP, São Paulo, SP. p. 183-189, 1988.

DIAS, M. *Absenteísmo em Contact Center*: Estudo de caso com preditores micro e macro-organizacionais. Dissertação de Mestrado em Psicologia Aplicada – Universidade Federal de Uberlândia, MG. 112 p., 2005.

DOU - Diário Oficial da República Federativa do Brasil. BRASIL. Decreto nº. 91.030, de 21 de janeiro de 1985. Disciplina o Despacho Aduaneiro de Importação e de Exportação de Remessas Expressas. Disponível em:

< <http://www.receita.fazenda.gov.br/legislacao/ins/2002/in1222002.htm>>. Acesso em: 01 fev. 2012.

DOUGHERTY, J.; KOHAVI, R.; SAHAMI, M. *Supervised and Unsupervised Discretization of Continuous Features*. Proc. 12 th Int. Conf. Machine Learning, 194-202, 1995.

FARIA, E. L.; ALBUQUERQUE, M. P.; ALFONSO, J. L. G.; ALBUQUERQUE, M. P.; CAVALCANTE, J. T. P. Introdução ao *Toolbox* de Redes Neurais de Kohonen. Centro Brasileiro de Pesquisas Físicas – CBPF. Ministério da Ciência Tecnológica – Governo Federal, 2010.

FAVA, Luiz Roberto. Absenteísmo e Presenteísmo. O seu Portal de Gestão. Carreira/qualidade de vida. São Paulo. Disponível em:

<http://www.ogerente.com.br/novo/colunas_ler.php?canal=6&canallocal=53&canalsub2=177&id=839> Acesso em: 21 nov. 2011.

FAYYAD, U. M.; 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, 1996a.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. UTHURUSAMY R. *Advances in Knowledge discovery and Data Mining*. MIT Press, 1996b.

- FERREIRA, R. P. Combinação de Técnicas da Inteligência Artificial para Previsão do Comportamento do Tráfego Veicular Urbano na Cidade de São Paulo. Dissertação de Mestrado em Engenharia de Produção – Universidade Nove de Julho. 107 p. São Paulo, 2011.
- FIORELLI, J. O. Psicologia para administradores: integrando teoria e prática 5ª. ed. São Paulo, SP: Atlas, 325 p., 2008.
- FOINA, P. R. Tecnologia de Informação: Planejamento e Gestão. 2ª. ed. São Paulo, SP: Atlas, 339 p., 2006.
- FONSECA, V. R. Atividade Física, Absenteísmo e Demanda por Atendimento à Saúde de Funcionários de Indústria Automobilística de São Caetano do Sul. 2009. Tese de Doutorado em Ciências – Universidade de São Paulo. 154 p., São Paulo, 2009.
- FREITAS, A. A. *A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery*. Curitiba, 2003.
- FREITAS, A. L. P.; SOUZA, R. G. B. Um Modelo para Avaliação da Qualidade de Vida no Trabalho em Universidades Públicas. *Revista Eletrônica Sistemas & Gestão* 4 (2): p. 136-154. Programa de Pós-Graduação em Sistemas de Gestão, Engenharia de Produção. Campos dos Goytacazes, Rio de Janeiro, TEP/TCE/CTC/PROPP/UFF, 2009.
- GARCIA, F. C.; SILVA, M F. G. Causas do Absenteísmo nas Organizações: um estudo de caso em Unidades de Alimentação e Nutrição. XII SemeAd – Empreendedorismo e Inovação. FEA/USP, São Paulo, 2009.
- GEHRING JUNIOR, G. CORRÊA FILHO, H. R. VIEIRA NETO, J. D. FERREIRA, N. A. VIEIRA, S.V. R. Absenteísmo-doença entre profissionais de enfermagem da rede básica do SUS Campinas. *Revista brasileira de epidemiologia*. vol.10 n.3, São Paulo, set. 2007.
- GIL, A. C. Como Elaborar Projetos de Pesquisa. 4ª. ED. São Paulo: ATLAS, 2002.
- GOETZEL, R. Z.; OZMINKOWSKI, R. J.; SEDERER, L.I.; MARRK, T.L. *The Business Case for Quality Mental Health Services: Why Employers Should Care About the Mental Health and Well-Being of Their Employees*. *Journal Of Occupational And Environmental Medicine*. v.44. p. 320-330, 2002.
- GOETZEL, R. Z.; LONG, S. R; OZMINKOWSKI, R. J; HAWKINS, K.; WANG, S. H; LYNCH, W. H. *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, 2004.
- GOLDSCHMIDT, R.; PASSOS, E. *Data Mining um guia prático*. Conceitos, Técnicas, Ferramentas, Orientações e Aplicações. Ed. Campus, Rio de Janeiro, 2005.

GROSSBERG, S. *How does a brain build a cognitive code?* *Psychological Review*. vol. 87, p. 1-51, 1980.

GUELPELI, M. V. C. Cassiopeia: Um Modelo Baseado em Sumarização e Aprendizado Autônomo Usado em Agrupamentos para Descoberta de Conhecimento em Bases textuais. Tese de Doutorado em Computação – Universidade Federal Fluminens. 100 p. Niterói, RJ, 2009.

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

HEBB, D. O. *The Organization of Behavior: Neuropsychological Theory*, N. Y.: Willey, 1949.

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.

HITO, S. C. Condições de Trabalho e Absenteísmo Odontológico em uma Indústria Frigorífica no Brasil: uma contribuição à saúde Ocupacional. Dissertação de Mestrado em Gestão Integrada em Saúde do Trabalho e Meio Ambiente – Centro Universário Senac. 130 p. São Paulo, SP, 2007.

HOPFIELD, J. J. *Neural network and phisical systems with emergent collective computational abilities*. *Proceedings of the Academy of Sciences of the USA*, vol.79, n.8, p. 2554-2558, 1982.

INOUE, K. C.; MATSUDA, L. M.; SILVA, D. M. P. P.; UCHIMURA, T. T.; MATHIAS, T. A. F. Absenteísmo-doença da equipe de enfermagem em unidade de terapia intensiva. *Rev. bras. enfermagem*. vol.61 nº.2. Mar./Apr., Brasília, 2008.

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

JOHNS, G. *How methodological diversity has improved our understanding of absenteeism from work*. *Human Resource Management Review*, 13 (2): p. 157-184, 2003.

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

JUNKES, M. B. Ônus do Absenteísmo de Médicos e Profissionais de Enfermagem que atuam em Hospitais Públicos da Região Centro Sul do Estado de Rondônia – Brasil. Tese de Doutorado em Ciências da Saúde – Univrsidade de Brasília.128 p. BSB, 2010.

KASKI, S.; KOHONEN, T. *Winner-Takes-All Networks*. *Triennial Report 1994 – 1996*, *Neural Networks Research Centre & Laboratory of Computer and Information Science*, *Helsinki University of Technology*, Finland, p. 72-75, 1997.

- KIVILUOTO, K. *Topology Preservation in Self-Organizing Maps. Technical Report A 29*, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1995.
- KOHONEN, T. *Self-organized formation of topologically correct feature maps. Biological Cybernetics*, vol. 43, p. 59-69, 1982.
- KOHONEN, T. *Self organizing maps for large documents, IEEE Transactions on Neural Networks, (Special Issue on Data Mining)*, vol. 11, p. 574–589, 2000.
- KOHONEN, T. *Self-Organizing Maps*. New York. Springer. 3^a ed., 2001.
- KOVÁCS, Z. L. *Redes Neurais Artificiais - Fundamentos e Aplicações*. Editora Livraria da Física. 4^a ed. São Paulo, 2006.
- LACANNA, M. L. *Qualidade de Vida no Trabalho*. Associação Brasileira de Qualidade de Vida. Vitale. São Paulo, 2008.
- LARANJEIRA, C. A. O Contexto Organizacional e a Experiência de *Estress*: uma Perspectiva Integrativa. *Revista salud pública* [online], vol.11, n.1, p. 123-133., 2009.
- LEVIN-EPSTEIN, J. *Presenteeism and Paid Sick Days*. Center for LAW and Social Policy – CLASP, ago. 2005.
- LIKERT, R. *A Technique for the Measurement of Attitudes. Archives of Psychology* 140, p. 1-55, 1932.
- LOPES, N.; RIBEIRO, B. *A Strategy for Dealing With Missing Values by Using Selective Activation Neurons in a Multi-Topology Framework. IEEE World congress on Computational Intelligence WCCI*, 2010.
- LOPES, N.; RIBEIRO, B. *GPU Implementation of the Multiple Back-Propagation Algorithm, In Proceedings of Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, v. 5788, p. 449-456, 2009.
- LOZANO, S.; ESPARZA, J.; ADENSO-DÍAZ, B.; GARCÍA, J. M. *Clustering Spanish households E-waste disposal behavior using Self-Organizing Feature Maps. IEEE Proceedings of Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference*. p. 2328 – 2332, 2010.
- LUDWIG JUNIOR, O.; MONTGOMERY, E. *Redes Neurais: Fundamentos e Aplicações com Programas em C*. Rio de Janeiro, RJ: ed. Ciência Moderna Ltda, 2007.

- LUFT, C. D. B.; SANCHES, S. O.; MAZO, G. Z.; ANDRADE, A. Versão Brasileira da Escala de Estresse Percebido: Tradução e Validação para idosos. *Rev. Saúde Pública*; 41(4):606-15, 2007.
- LUMARE JÚNIOR, G. Valor econômico do cliente: uma teoria das encomendas. São Paulo: *Financial Times – Prentice Hall*, 2007.
- MAASS, W.; BISHOP, C. M. *Pulsed Neural Networks*. Cambridge: MIT Press, 2001.
- MALLADA, F. J. R. A gestão do absentismo trabalhista nas empresas espanholas. Disponível em:
<http://www.kplus.com.br/materia.asp?co=131&rv=Direito> Acesso em: 15 jan. 2012.
- MARRAS, J. P. Administração de recursos humanos: do operacional ao estratégico. São Paulo: Futura, 2000.
- MARTEL, J. P.; DUPUIS, G. *Quality of work life: theoretical and methodological problems, and presentation of a new model and measuring instrument*. *Social Indicators Research*, vol. 77, p. 333 – 368, 2006.
- MARTINEZ, L. F.; FERREIRA, A. I.; SOUZA, L. M.; CUNHA, J. V. A Esperança é a última a morrer? Capital Psicológico Positivo e Presentismo. *Comportamento Organizacional e Gestão*, vol. 13, n. 1, p. 37-54, 2007.
- MARTINS, P. L.; SILVA, E. M. S.; ROCHA, R. C. Fatores Prejudiciais ao bom desempenho dos Funcionários nas Diversas Organizações: com Ênfase para o Alcoolismo. VI SEGeT – Simpósio de Excelência em Gestão e Tecnologia, 2009.
- MATTKE, S.; BALAKRISHNAN, A.; BERGAMO, G.; NEWBERRY, S. J. *A Review of Methods to Measure Health-related Productivity Loss*. *The American Journal of Managed Care*. vol. 13, n. 4. p. 211-217. *Review*, 2007.
- MCCULLOCH, W. S.; PITTS, W. H. *A logical calculus of the ideas immanent in nervous activity*, *Bulletin of mathematical biophysics*, vol. 5, p. 115-133, 1943.
- McGEE, J. V.; PRUSAK, L. Gerenciamento Estratégico da Informação: Aumente a competitividade de sua empresa utilizando a informação como uma ferramenta estratégica. 17ª ed. Rio de Janeiro: Elsevier, 244 p., 1994.
- MCKEVITT TAUFFER, R.; COLTRE, S. M. 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. 2007.
- MINSKY, M. L.; PAPER, S. A. *Perceptrons: An Introduction to Computational Geometry, expanded edition*, Cambridge, MA: MIT Press, 1969.

MIRANDA, A. R.; GOMES, C. F.; GOMES, M. A. N. Uma avaliação da Qualidade de Vida no Trabalho dos gerentes da construção civil. XIII SIMPEP, Bauru – SP, Brasil, nov. de 2006.

MITCHELL, T. M. *Machine Learning*. McGraw-Hill, 1997.

MURTA, S. G.; LAROS, J. A.; TRÓCCOLI, B. T. Manejo de estresse ocupacional na perspectiva da área de avaliação de programas. Estudos de Psicologia. Natal. vol.10 n.2 Natal mai/ago. p.167-176, 2005.

NASCIMENTO JÚNIOR, C. L.; YONEYAMA, T. Inteligência Artificial em Controle e Automação. São Paulo: ed. Blucher: FAPESP, São Paulo, 2004.

NG, T. S.; HUANG, H. C. NG, J. Y. *Human Resource Planning with Worker Attendance Uncertaint. Engineering and Engineering Management. IEEE International Conference on., Singapore, 2008.*

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

O' DONNELL, J. E. *Presenteeism: A Comparative Analysis. Ph.D. thesis – University of Massachusetts, Amherst, 2009.*

PACHECO, M. A. C.; VELLASCO, M. M. B. R. (org) Sistemas inteligentes de apoio à decisão: análise econômica de projetos de desenvolvimento de campos de petróleo sob incerteza. Rio de Janeiro: Ed. PUC-Rio: Ed. Interciência, 2007.

PAFARO, R. C.; MARTINO, M. M. F. Estudo do estresse do enfermeiro com dupla jornada de trabalho em um hospital de oncologia pediátrica de Campinas. Revista de Escola de enfermagem da USP. vol.38 n.2. São Paulo, jun. 2004.

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

PAUGAM-MOISY, H.; BOHTE, S. *Computing with spiking neuron networks. In: Rozenberg, G., Back, T., Kok, J. (Eds.), Handbook of Natural Computing. Springer-Verlag, 2010.*

PAULY, M. Efeito multiplicador: conseqüências financeiras do absenteísmo. Disponível em: <http://www.wharton.universia.net/index.cfm?fa=viewfeature&id=1075&language=portuguese>
Acesso em: 04 jan. 2012.

PENATTI, I.; ZAGO, J.S.; QUELHAS, O. Absenteísmo: As conseqüências na gestão de pessoas. III SEGeT – Simpósio de Excelência em Gestão e Tecnologia, 2006.

PIATETSKY-SHAPIO, G.; MATHEUS, C. J.; CHAN, P. K. *Systems for Knowledge Discovery in Data bases*. IEEE,1993, p. 903–912, 1993.

PRIMO, G. M. G. O Perfil dos Trabalhadores, seu Adoecimento e Absenteísmo em um Hospital Público Universitário. Dissertação de Mestrado em Saúde Pública na área de concentração em saúde e trabalho – Universidade Federal de Minas Gerais. 110 p. MG, 2008.

ROCHA, P. R. F. O Trabalhador-Doente e sua Família nos seus Papéis como Atores Sociais no Processo de Retorno ao Trabalho. Dissertação de Mestrado em Medicina na área da Saúde, Ambiente e Trabalho – Universidade Federal da Bahia. 120 p. BA, 2010.

RODRIGUES, J. A . F. *Data Mining: conceitos, técnicas e aplicação*. Dissertação de Mestrado em Engenharia de Produção - Escola Politécnica da Universidade de São Paulo, SP - EPUSP, São Paulo, 2001.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. *Learning Internal Representations by Error Propagation*. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations, D. E. Rumelhart and J. L. McClelland, Eds. Mit Press *Computational Models Of Cognition And Perception Series*. MIT Press, Cambridge, MA, p. 318-362, 1986.

SALLES, J. A. A; IOZZI, L. O. Contribuições para a configuração de um sistema de medição de desempenho para incubadoras de empresas baseado no BSC. *Revista Exacta*, v. 8, n. 2, p. 145-156, São Paulo, 2010.

SANTOS, L. A. Atividade física e morbidade cardiovascular referidas pelos gerentes e diretores de uma indústria automobilística: influência de um programa de condicionamento físico supervisionado. Tese de Doutorado em Ciências – Universidade de São Paulo, São Paulo. 144 p. SP, 2008.

SASSI, R. J. *An hybrid architecture for clusters analysis: rough setstheory and self-organizing map artificial neural network*. *Pesquisa Operacional*, Apr 2012, v. 32, n.1, p.139-164, 2012.

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. Tese de Doutorado em Engenharia Elétrica – Escola Politécnica da Universidade de São Paulo. 169 p. SP, 2006.

SASSI, R. J.; SILVA, L. A. ; HERNANDEZ, E. del 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, out. 2008.*

SELAU, L. P. R.; RIBEIRO, J. L. D. Uma sistemática para construção e escolha de modelos de previsão de risco de crédito. *Gestão Produção. v. 16, n. 3, p. 398-413, jul/set. São Carlos, SP, 2009.*

SILVA, D. M.P.P.da; MARZIALE, M.H.P. Absenteísmo de Trabalhadores de Enfermagem em um Hospital Universitário. *Net, Ribeirão Preto. Rev. Latino-Am. Enfermagem vol.8 n.5. Disponível em:*

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-11692000000500007>.

Acesso em: 11 jan. 2012a.

SILVA, D. M. P. P.; MARZIALE, M. H.P. Problemas de Saúde Responsáveis pelo Absenteísmo de Trabalhadores de Enfermagem em um Hospital Universitário. *Acta Scientiarum. Health Sciences. DOI: 10.4025/actascihealthsci. V.25i2.2232, vol. 25 n. 2, p. 191-197, Maringá. Disponível em:*

<http://periodicos.uem.br/ojs/index.php/ActaSciHealthSci/article/viewFile/2232/1459>>.

Acesso em: 06 jan. 2012b.

SILVA, E. L.; MENEZES E. M. *Metodologia da pesquisa e elaboração de dissertação. 3.ed, Florianópolis, 2001.*

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

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

SILVA NETO, M. A. *Mineração Visual de Dados: extração do conhecimento a partir das técnicas de Visualização da Informação em mineração de dados. Dissertação de Mestrado em Ciências na área de concentração Programação Matemática – Universidade Federal do Paraná. 172 p. PR, 2008.*

SIMÕES, M. G.; SHAW, I. S. *Controle e Modelagem fuzzy. São Paulo: Blucher: FAPESP, 2007.*

SVENSÉN, J.F.M. *GTM: The Generative Topographic Mapping*. PhD Thesis, Aston University, April 1998.

TAUFFER, R.; COLTRE, S. M. A Contribuição dos Fatores Intrínsecos e Extrínsecos da Organização para o Absenteísmo nos Setores de Produção de uma Indústria de Carrocerias de Ônibus. XXVII Encontro Nacional de Engenharia de Produção – ENEGEP. Foz de Iguaçu, PR. p.1-9, 2007.

THORWARTH, M. ARISHA, A.; HARPER, P. *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, 2009.

ULTSCH, A.; SIEMON, H.P. *Kohonen's Self Organizing Feature Maps for exploratory Data Analysis*. *Proceedings of the International Neural Networks Conference*. Dordrecht. Netherlands, p. 305-308, 1989.

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

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

VESANTO, J.; ALHONIEMI, E. *Clustering of the Self-Organizing Map*. *IEEE Transactions on Neural Networks*, vol. 11, n. 2, p. 586-600, 2000.

VESANTO, J. *Data Exploration Process Based on the Self-Organizing Map*. PhD thesis, Helsinki University of Technology, 2002.

WU, Z.; WENG, M. X. *Multiagent Scheduling Method With Earliness and Tardiness Objectives in Flexible Job Shops* *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, v.35, n. 2, p. 293-301, 2005.

ZHUANG, Z. Y.; CHURILOV, L.; SIKARIS, K. *Uncovering the Patterns in Pathology Ordering by Australian General Practitioners. A Data Mining Perspective*. *System Sciences, 2006. HICSS'06. IEEE Proceedings of the 39th Annual Hawaii International Conference*. v. 5, p. 92c – 92c., 2006.

APÊNDICE A – QUESTIONÁRIO

Este questionário procura avaliar os seus sentimentos e pensamentos durante os últimos anos com relação ao estresse. Em cada questão, será pedido para você indicar o quanto frequentemente tem se sentido com relação ao estresse percebido.

Deve-se responder a cada questão rapidamente, ou seja, não avalie o número de vezes que se sentiu de uma maneira particular, mas indique a alternativa que lhe parece razoável. Para cada pergunta, escolha uma entre as seguintes alternativas:

- a) 0 = nunca
- b) 1 = quase nunca
- c) 2 = às vezes
- d) 3 = quase sempre
- e) 4 = sempre

Tabela A – Questionário para identificar o nível de estresse percebido

Questões	Nos últimos anos, com que frequência...
1	Você tem ficado triste por causa de algo que aconteceu inesperadamente?
2	Você tem se sentido incapaz de controlar as coisas importantes em sua vida?
3	Você tem se sentido nervoso e “estressado”?
4	Você tem tratado com sucesso dos problemas difíceis da vida?
5	Você tem sentido que está lidando bem com as mudanças importantes que estão ocorrendo em sua vida?
6	Você tem se sentido confiante na sua habilidade de resolver problemas pessoais?
7	Você tem sentido que as coisas estão acontecendo de acordo com a sua vontade?
8	Você tem achado que não conseguiria lidar com todas as coisas que você tem que fazer?
9	Você tem conseguido controlar as irritações em sua vida?
10	Você tem sentido que as coisas estão sob o seu controle?
11	Você tem ficado irritado porque as coisas que acontecem estão fora do seu controle?
12	Você tem se encontrado pensando sobre as coisas que deve fazer?
13	Você tem conseguido controlar a maneira como gasta seu tempo?
14	Você tem sentido que as dificuldades se acumulam a ponto de você acreditar que não pode superá-las?

Fonte: Adaptado de Luft *et al.*, (2007).

APÊNDICE B – PUBLICAÇÕES DO AUTOR

Em Periódico Nacional:

1. Martiniano, A. S., SASSI, R. J, Vieira, A. B. Um Estudo sobre a Qualidade de Vida no Trabalho: O Absenteísmo e o Presenteísmo. INNOVARE. , v.II, p. 157 - 170, 2010.

Em Anais de Congressos Internacionais:

1. Martiniano, A. S., Ferreira, Affonso, C., R. P., SASSI, R. J. Aplicação de uma Rede Neuro Fuzzy na Previsão do Absenteísmo no Trabalho. CISTI2012 - 7ª Conferencia Ibérica de Sistemas y Tecnologías de Información, 2012.

2. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. *Control and Monitoring of the Indexes of Absenteeism and Presenteeism With aid of the Technology of the Information In: 7th CONTECSI - 2010, 2010, São Paulo. International Conference on Information Systems and Technology Management, 2010. v.7. p. 489 – 506.*

3. Martiniano, A. S., SASSI, R. J. Monitoramento do Presenteísmo através da Tecnologia da Informação In: SHEWC - 2010, 2010, São Paulo. Congresso Mundial de Pesquisas Ambientais, Saúde e Segurança. , 2010.

Em Anais de Congressos Nacionais:

1. Martiniano, A. ; Ferreira, R. P.; Sassi, R.J. Aplicação da Mineração de Dados na Identificação de Empregados Absenteístas e Presenteístas de uma Empresa de *Courier* da Cidade de São Paulo. VII – Encontro Mineiro de Engenharia de Produção - EMEPRO, Itajubá – MG, 2012.

2. Martiniano, A. S.; Ferreira, R. P.; SASSI, R. J. O Uso de Mapas Auto-Organizáveis de Kohonen Para Identificar e Prevenir o Absenteísmo. *In: XVIII Simpósio de Engenharia de Produção - SIMPEP, 2011, Bauru. Sustentabilidade na Cadeia de Suprimentos, 2011. v. XVIII. p. 1-12.*

3. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. Determinação do Perfil de Colaboradores Presenteístas: utilizando Técnicas de *Data Mining* In: VI EMEPRO – Encontro Mineiro de Engenharia de Produção, 2010, Coronel Fabriciano - MG. O Engenheiro de Produção e o Mercado de Oportunidades: Trajetórias e Perspectivas Tecnológicas. , v.VI. p.1 – 10, 2010.
4. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. Um estudo sobre o Absenteísmo através do Código Internacional de Doenças utilizando *Data Mining* In: VII Encontro Mineiro de Engenharia de Produção - EMEPRO, 2011, São João Del Rei – MG, 2011.
5. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. A Qualidade de Vida no Trabalho versus Absenteísmo e Presenteísmo In: SPOLM - Simpósio de Pesquisa Operacional e Logística da Marinha, 2010, Rio de Janeiro. A Qualidade de Vida no Trabalho *Versus* Absenteísmo e Presenteísmo. Rio de Janeiro: PRIMIL, 2010.
6. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. Mapas Auto-Organizáveis de Kohonen no Processo de Descoberta de Conhecimento In: 12º Simpósio de Iniciação Científica e Tecnológica - FATEC, 2010, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo, 2010. v.II. p.35 – 35.
7. Martiniano, A. S., Ferreira, R. P. *Control and Monitoring of the indexes of Absenteeism and of Presences with aid of the Technology* In: XLI SBPO 2009 - Simpósio Brasileiro de Pesquisa Operacional, 2009, Porto Seguro / Bahia. XLI SBPO 2009 - Pesquisa Operacional na Gestão do Conhecimento. p. 3223 – 3225, 2009.

Co-orientação em Projeto de Iniciação Científica

Foi realizada co-orientação em um projeto de Iniciação Científica, intitulado: “Determinação do Perfil de Colaboradores Presenteístas: Utilizando Técnicas de Data Mining”.

Iniciado em 2010 e finalizado em 2011, este projeto de Iniciação Científica gerou os seguintes artigos publicados com os alunos de graduação: Anselmo Bernardes Vieira, Rogério Pereira da Silva, Thiago Correa Figueredo e Daniela das Dores de Fátima.

1. Martiniano, A. S., Ferreira, R. P., Vieira, A. B., Silva, R. P., SASSI, R. J. Aplicação do Data Mining para Identificação do Absenteísmo e Presenteísmo: 13º Simpósio de Iniciação Científica e Tecnológica - FATEC, 2011, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo, 2011. v.31. p. 25 – 25.

2. Martiniano, A. S., Ferreira, R. P., Vieira, A. B., Silva, R. P., SASSI, R. J. Mapas Auto-organizáveis de Kohonen para Identificar e Prevenir o Absenteísmo: 13º Simpósio de Iniciação Científica e Tecnológica - FATEC, 2011, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo, 2011. v.31. p. 96 – 96.
3. Martiniano, A. S., Ferreira, R. P., Affonso, C., Vieira, A. B., SASSI, R. J. *Application of a Neuro Fuzzy Network to Absenteeism Prediction from Work in a Company Courier In: XLIII Simpósio Brasileiro de Pesquisa Operacional - SBPO, 2011, Ubatuba-SP. XLIII Simpósio Brasileiro de Pesquisa Operacional - SBPO, 2011.*
4. Vieira, A. B.; Martiniano, A. S.; Ferreira, R. P.; SASSI, R. J. Redes Neurais Artificiais para identificar e prevenir o Absenteísmo In: SIICUSP - Simpósio Internacional de Iniciação Científica da Universidade de São Paulo, 2011, São Paulo. 19º SIICUSP - Simpósio Internacional de Iniciação Científica. v. XIX, 2011.
5. Fátima, D. D., Figueredo, T. C., Martiniano, A. S., Ferreira, R. P., SASSI, R. J. *The contribution to the Psychology of work for reduction of Absenteeism using Data Mining In: XLIII Simpósio Brasileiro de Pesquisa Operacional - SBPO, 2011, Ubatuba-SP - XLIII Simpósio Brasileiro de Pesquisa Operacional – SBPO, 2011.*
6. Vieira, A. B., Souza, E. P. C., Silva, R. P., Martiniano, A. S., SASSI, R. J. Um Estudo sobre a Qualidade de Vida no Trabalho o Absenteísmo e o Presenteísmo *In: 22º Congresso Internacional de Administração, 2010, Ponta Grossa/PR. Gestão Estratégica: Inovação, Tecnologia e Sustentabilidade. v. 22. p. 1 – 8, 2010.*
7. Figueredo, T. C., Martiniano, A. S., Ferreira, R. P., SASSI, R. J. Um estudo do Absenteísmo através do Código Internacional de Doenças *In: 12º Simpósio de Iniciação Científica e Tecnológica - FATEC, 2010, 2010, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo. v. 29. p. 61 – 61, 2010.*
8. Martiniano, A. S., Ferreira, R. P., SASSI, R. J. Mapas Auto-Organizáveis de Kohonen no Processo de Descoberta de Conhecimento In: 12º Simpósio de Iniciação Científica e Tecnológica - FATEC, 2010, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo, 2010. v.II. p.35 – 35.
9. Silva, R. P., Martiniano, A. S., SASSI, R. J. Um estudo do Absenteísmo numa empresa de *Courier* através da classificação do Código Internacional de Doenças *In: SIICUSP - Simpósio Internacional de Iniciação Científica da Universidade de São Paulo, 2010, São Paulo. 18º SIICUSP - Simpósio Internacional de Iniciação Científica. v. XVIII, 2010.*

10. Martiniano, A. S., Souza, E. P. C., Silva, R. P., Vieira, A. B., SASSI, R. J. *Using Self-Organizing Map – Som of Kohonen Discovery in Presenteeism Behavior in a Employee Database In: 42º SBPO - Simpósio Brasileiro de Pesquisa Operacional – SBPO. 2010, Bento Gonçalves. Saúde Humana, Saúde Animal e Ecosistema , 2010. v.42º.*