

**UNIVERSIDADE NOVE DE JULHO – UNINOVE
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA E GESTÃO
DO CONHECIMENTO**

ANDRÉA MARTINIANO DA SILVA

**PREVISÃO DO COMPORTAMENTO DOS EMPREGADOS NO TRABALHO:
MINERAÇÃO DE DADOS APLICADA EM BASE DE DADOS
ANTROPOMÉTRICOS, ERGONÔMICOS, ABSENTEÍSTAS E PRESENTEÍSTAS**

**São Paulo
2020**

ANDRÉA MARTINIANO DA SILVA

**PREVISÃO DO COMPORTAMENTO DOS EMPREGADOS NO TRABALHO:
MINERAÇÃO DE DADOS APLICADA EM BASE DE DADOS
ANTROPOMÉTRICOS, ERGONÔMICOS, ABSENTEÍSTAS E PRESENTEÍSTAS**

Tese de doutorado apresentada ao Programa de Pós-graduação em Informática e Gestão do Conhecimento – PPGI da Universidade Nove de Julho, como requisito parcial para obtenção do título de Doutora em Informática e Gestão do Conhecimento.

Orientador: Prof. Dr. Renato José Sassi

São Paulo
2020

Silva, Andréa Martiniano da.

Previsão do comportamento dos empregados no trabalho: mineração de dados aplicada em base de dados antropométricos, ergonômicos, absenteístas e presenteístas. / Andréa Martiniano da Silva. 2020.

141 f.

Tese (Doutorado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2020.

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

1. Mineração de dados. 2. Redes neurais artificiais. 3. Antropometria. 4. Ergonomia. 5. Absenteísmo.

Sassi, Renato José.

II. Título.

CDU 004

Dedico este trabalho aos meus pais Jurandir Martiniano da Silva e Tereza Corrêa da Silva, e de maneira muito especial ao meu esposo Ricardo Pinto Ferreira.

“Nós só podemos ver um pouco do futuro, mas
o suficiente para perceber que há muito a fazer”
Alan Turing

AGRADECIMENTOS

A Deus, o “Maravilhoso Conselheiro”, que me ilumina, me fortalece e me consola nos momentos mais difíceis. Devo a Ele tudo o que tenho recebido e tudo o que sou.

Aos meus pais Jurandir Martiniano da Silva e Tereza Corrêa da Silva, pela sabedoria do sempre presente estímulo ao estudo e ao crescimento contínuo, por todo amor que me dedicaram.

De maneira muito especial, ao meu esposo Ricardo que partilha comigo cada momento em minha vida, sempre me deu forças para seguir em frente em busca de conhecimento e ao apoio e dedicação em toda etapa deste trabalho.

Aos amigos que contribuíram, direta e indiretamente para que este trabalho fosse realizado, obrigada pelo carinho e amizade.

Ao meu Orientador Prof. Dr. Renato José Sassi, pela participação no desenvolvimento deste trabalho pelo incentivo, disponibilidade e paciência em compartilhar seu conhecimento, que certamente contribuíram para meu aprendizado.

Aos docentes e funcionários da Universidade Nove de Julho, pela atenção e profissionalismo, em especial, ao Prof. Dr. André Felipe Henriques Librantz.

Aos membros da banca, Prof. Dr. Marco Antonio Silveira, Prof. Dr. Fabio Silva Lopes, Prof. Dr. André Felipe Henriques Librantz, Prof. Dr. Felipe Silva Martins e Prof. Dr. Domingos Márcio Napolinato pelo pronto atendimento ao convite.

A todos os alunos do Programa de Pós-Graduação em Informática e Gestão do Conhecimento - PPGI da Universidade Nove de Julho, em especial, aos orientados do Prof. Dr. Renato José Sassi.

À Paloma Blanca Alves Barbieri, pela revisão na língua vernácula.

À Universidade Nove de Julho pela bolsa e apoio à pesquisa e à PROSUP/CAPES, pelo auxílio financeiro que me foi concedido durante o desenvolvimento deste trabalho.

Aos meus familiares e aos meus irmãos André Martiniano da Silva, Rogério Martiniano da Silva e Andreza Martiniano da Silva, que contribuíram para o meu sucesso, de maneira muito especial à minha sobrinha e afilhada Sarah Martiniano Soledade de Oliveira.

RESUMO

O absenteísmo é um fenômeno definido como o não comparecimento do empregado ao local de trabalho de forma habitual e com frequência regular; por conseguinte, é o não cumprimento das obrigações, conforme o programado. O presenteísmo, por outro lado, é o fenômeno que indica a presença do empregado, ainda que doente, no local de trabalho, porém, a realização de suas atividades e de suas funções pode ocorrer de modo improdutivo. Neste sentido, dados antropométricos e ergonômicos, que são medidas corporais, mostram-se importantes quando relacionados ao absenteísmo e ao presenteísmo, principalmente em atividades classificadas como trabalho pesado e com um alto índice de atividades repetitivas. A previsão do comportamento de empregados é importante para reduzir perdas para a empresa e melhorar a qualidade de vida no trabalho. Diante deste cenário, técnicas de Mineração de Dados e algumas áreas da Inteligência Artificial podem ser aplicadas na previsão do comportamento do empregado no trabalho. Assim, o objetivo deste estudo foi investigar a aplicação da mineração de dados, em base de dados antropométricos, ergonômicos, absenteístas e presenteístas, para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no ambiente de trabalho. Os experimentos computacionais foram desenvolvidos em três etapas a fim de prever o comportamento dos empregados – comportamento este que pode ser classificado como presenteísta, normal e absenteísta. Para realização dos experimentos, duas arquiteturas diferentes de redes neurais artificiais foram aplicadas: a *Multilayer Perceptron* (MLP) e a *Radial Basis Function* (RBF). Ademais, utilizaram-se também a *Random Forest* (RF) e o Algoritmo de Otimização por Colônia de Formigas, *Ant Colony Optimization* (ACO). Para enriquecer o experimento, três bases de dados foram utilizadas. Os dados de absenteísmo e presenteísmo são comuns às três bases de dados e são compostas com 2.403 registros de licenças médicas de 39 empregados coletados durante o período de janeiro de 2008 a dezembro de 2017. Foram considerados 10 atributos na primeira base, 11, na segunda, e 25, na terceira base. Com exceção da primeira base, a segunda e a terceira foram enriquecidas com dados antropométricos e ergonômicos. Os resultados mostraram melhor desempenho após o enriquecimento das bases de dados por meio da MLP e da rede SOM. Em relação aos resultados dos experimentos computacionais na Etapa 1, a MLP obteve a taxa de acerto de 99,91%, a RBF, de 97,08%, a RF, de 99,91%, e o ACO, de 80,65%. Na Etapa 2, a MLP obteve a taxa de acerto de 99,91%, a RBF, de 97,25%, a RF, de 99,91%, e o ACO, de 84,44%. Na Etapa 3, a MLP obteve a taxa de acerto de 99,96%, a RBF, de 96,25%, a RF, de 99,91%, e o ACO, de 91,80%. Em relação ao tempo de processamento e ao desempenho, a RF se destacou como sendo a técnica mais recomendada para auxiliar na previsão dos comportamentos presenteísta, normal e absenteísta, no ambiente de trabalho.

Palavras-chave: Mineração de Dados, Redes Neurais Artificiais, Antropometria, Ergonomia, Absenteísmo, Presenteísmo.

ABSTRACT

Absenteeism is a phenomenon defined as the employee's failure to show up to the workplace in a regular and regular manner; therefore, it is the non-fulfillment of obligations, as scheduled. Presenteeism, on the other hand, indicates the presence of the employee, albeit ill, in the workplace, however, the performance of his activities and functions may occur in an unproductive way. In this sense, anthropometric and ergonomic data, which are body measurements, are important when related to absenteeism and presenteeism, especially in activities classified as heavy work and with a high rate of repetitive activities. Predicting employee behavior is important to reduce losses for the company and improve the quality of life at work. Given this scenario, Data Mining techniques and some areas of Artificial Intelligence can be applied in predicting employee behavior at work. Thus, the objective of this study was to investigate the application of data mining, based on anthropometric, ergonomic, absenteeist and presenteeist data, to help predict the behavior of employees in the work environment. The computational experiments were developed in three stages in order to predict the behavior of employees - behavior that can be classified as presenteeist, normal and absenteeist. To carry out the experiments, two different architectures of artificial neural networks were applied: the Multilayer Perceptron (MLP) and the Radial Basis Function (RBF). In addition, Random Forest (RF) and Ant Colony Optimization (ACO) were also used. To enrich the experiment, three databases were used. Absenteeism and presenteeism data are common to the three databases and consist of 2,403 medical leave records from 39 employees collected during the period from January 2008 to December 2017. 10 attributes were considered in the first base, 11 in the second, and 25, on the third base. With the exception of the first base, the second and third were enriched with anthropometric and ergonomic data. The results showed better performance after the enrichment of the databases through MLP and the SOM network. In relation to the results of the computational experiments in Step 1, MLP obtained a hit rate of 99.91%, RBF, 97.08%, RF, 99.91%, and ACO, 80.65 %. In Step 2, the MLP obtained a hit rate of 99.91%, the RBF, 97.25%, the RF, 99.91%, and the ACO, 84.44%. In Step 3, the MLP obtained a hit rate of 99.96%, the RBF, 96.25%, the RF, 99.91%, and the ACO, 91.80%. Regarding processing time and performance, RF stood out as being the most recommended technique to assist in the prediction of presenteeist, normal and absenteeist behaviors in the work environment.

Keywords: Data Mining, Artificial Neural Networks, Anthropometry, Ergonomics, Absenteeism, Presenteeism.

SUMÁRIO

1. INTRODUÇÃO	1
1.1 IDENTIFICAÇÃO DE LACUNA DE PESQUISA.....	4
1.2 JUSTIFICATIVA E MOTIVAÇÃO.....	6
1.3 PROBLEMA DE PESQUISA.....	7
1.4 OBJETIVOS.....	8
1.4.1 OBJETIVO GERAL.....	8
1.4.2 Objetivos específicos.....	8
1.5 DELIMITAÇÃO DO TEMA.....	9
1.6 ORGANIZAÇÃO DO TRABALHO.....	9
2. FUNDAMENTAÇÃO TEÓRICA.....	10
2.1 ABSENTEÍSMO.....	10
2.1.1 O absenteísmo nas empresas.....	11
2.1.2 Pesquisas relacionadas ao Absenteísmo.....	12
2.2 PRESENTEÍSMO.....	15
2.2.1 O presenteísmo nas empresas.....	16
2.2.2 Pesquisas relacionadas ao Presenteísmo.....	17
2.3 ANTROPOMETRIA.....	19
2.3.1 Pesquisas relacionadas à antropometria.....	21
2.4 ERGONOMIA.....	23
2.4.1 Medições em ergonomia.....	24
2.4.2 Especialização da ergonomia.....	25
2.4.3 Pesquisas relacionadas a ergonomia.....	26
2.5 MINERAÇÃO DE DADOS.....	29
2.5.1 Tarefas de mineração de dados.....	31
2.5.2 Pesquisas relacionadas à mineração de dados.....	34
2.6 TÉCNICAS DE MINERAÇÃO DE DADOS	37
2.6.1 Redes Neurais Artificiais (RNAs).....	37
2.6.1.1 Mapa Auto-Organizável de Kohonen (SOM).....	39
2.6.1.2 <i>Multilayer Perceptron</i> (MLP).....	40
2.6.1.3 <i>Radial Basis Function</i> (RBF).....	42
2.6.2 <i>Random Forest</i> (RF).....	43
2.6.3 Algoritmo de Otimização por Colônia de Formigas (ACO).....	45

3. MATERIAIS E MÉTODOS.....	49
3.1 CARACTERIZAÇÃO METODOLÓGICA.....	49
3.2 PESQUISA BIBLIOGRÁFICA.....	49
3.3 BASE DE DADOS E PLATAFORMA DE EXPERIMENTOS.....	56
3.3.1 Características dos atributos das bases de dados selecionadas.....	62
3.3.2 Descrição do enriquecimento das bases de dados.....	64
3.3.3 Bases de dados utilizadas para auxiliar na previsão do comportamento no trabalho.....	67
3.4 CONDUÇÃO DOS EXPERIMENTOS COMPUTACIONAIS.....	69
3.4.1 Parâmetros utilizados nas técnicas aplicadas.....	70
3.4.2 Estratégia de treinamento: validação cruzada (<i>k-fold cross validation</i>).....	70
3.4.3 Medidas de avaliação do desempenho.....	71
4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS.....	75
4.1 EXPERIMENTOS COMPUTACIONAIS REALIZADOS PARA O ENRIQUECIMENTO DAS BASES DE DADOS.....	75
4.2 REALIZAÇÃO DOS EXPERIMENTOS COMPUTACIONAIS.....	78
4.2.1 Experimentos computacionais – Etapa 1 – Base de dados Número 1 com 10 atributos.....	79
4.2.2 Experimentos computacionais – Etapa 2 – Base de dados Número 2 com 11 atributos.....	83
4.2.3 Experimentos computacionais – Etapa 3 – Base de dados Número 3 com 25 atributos.....	87
4.3 ANÁLISE DOS RESULTADOS DOS EXPERIMENTOS COMPUTACIONAIS.....	91
4.3.1 Análise dos resultados computacionais na Etapa 1.....	92
4.3.2 Análise dos resultados computacionais na Etapa 2.....	93
4.3.3 Análise dos resultados computacionais na Etapa 3.....	96
4.4 CONSOLIDAÇÃO DOS RESULTADOS COMPUTACIONAIS DAS ETAPAS 1,2 e 3.....	98
5. CONCLUSÃO.....	105
5.1 MEMORIAL DE PUBLICAÇÕES.....	108

5.2 BASES DE DADOS CEDIDAS PARA UCI <i>MACHINE LEARNING REPOSITORY</i>	109
5.3 REGISTRO DE PROGRAMA DE COMPUTADOR ORIUNDO DA PESQUISA.....	110
REFERÊNCIAS BIBLIOGRÁFICAS.....	111
Anexo A – Certificado de Registro de Programa de Computador.....	124

LISTA DE FIGURAS

Figura 1 – Etapas do processo simplificado de mineração de dados.....30

Figura 2 – Tarefa de classificação.....31

Figura 3 – Tarefa de agrupamento.....33

Figura 4 – Tarefa de associação.....34

Figura 5 – Mapa topológico unidimensional e bidimensional.....39

Figura 6 – Rede MLP.....40

Figura 7 – Configuração típica de rede RBF.....42

Figura 8 – Modelo de aprendizado por conjunto da RF.....43

Figura 9 – Probabilidade de seleção de cada destino.....46

Figura 10 – Seleção do caminho das formigas usando o feromônio.....47

Figura 11– Procedimentos para o enriquecimento da base de dados.....58

Figura 12 – Topologia da MLP 1.....66

Figura 13 – Topologia da MLP 2.....66

Figura 14 – Etapas dos experimentos computacionais.....69

Figura 15 – Exemplo de validação cruzada usando N partições.....71

Figura 16 – Agrupamentos gerados pela rede SOM.....76

Figura 17 – Topologia da MLP 1 utilizada no experimento.....77

Figura 18 – Topologia da MLP 2 utilizada no experimento.....78

Figura 19 – Quantidade de empregados com comportamentos presenteísta, normal e
absenteísta.....79

Figura 20 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP
na Etapa 1.....80

Figura 21 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RBF
na Etapa 1.....81

Figura 22 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF
na Etapa 1.....81

Figura 23 – Matriz de confusão com as classificações corretas e incorretas da aplicação do
ACO na Etapa 1.....82

Figura 24 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP
na Etapa 2.....83

Figura 25 – – Matriz de confusão com as classificações corretas e incorretas da aplicação da
RBF na Etapa 2.....84

Figura 26 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF na Etapa 2.....	85
Figura 27 – Matriz de confusão com as classificações corretas e incorretas da aplicação do ACO na Etapa 2.....	85
Figura 28 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP na Etapa 3.....	88
Figura 29 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RBF na Etapa 3.....	89
Figura 30 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF na Etapa 3.....	89
Figura 31 – Matriz de confusão com as classificações corretas e incorretas da aplicação do ACO na Etapa 3.....	90
Figura 32 – Taxa de acerto das quatro técnicas aplicadas para auxiliar a previsão do comportamento dos empregados.....	100
Figura 33 – Taxa de acerto e tempo de processamento das técnicas RBF e RF aplicadas para auxiliar a previsão do comportamento dos empregados.....	101
Figura 34 – Taxa de acerto e tempo de processamento das técnicas ACO e MLP aplicadas para auxiliar a previsão do comportamento dos empregados.....	103

LISTA DE TABELAS

Tabela 1 – Relação dos 30 artigos selecionados na pesquisa bibliográfica.....	51
Tabela 2 – Classificação Internacional de Doenças (CID).....	61
Tabela 3 – Lista dos seis motivos (categorias) sem CID.....	62
Tabela 4 – Exemplo de uma matriz de confusão.....	72
Tabela 5 – Interpretação dos valores do índice Kappa.....	73
Tabela 6 – Parâmetros utilizados na MLP 1 (dados de ergonomia)	76
Tabela 7 – Parâmetros utilizados na MLP 2 (dados antropométricos).....	77
Tabela 8 – Parâmetros utilizados na MLP aplicada na Etapa 1.....	79
Tabela 9 – Parâmetros utilizados na RBF aplicada na Etapa 1.....	80
Tabela 10 – Parâmetros utilizados na MLP aplicada na Etapa 2.....	83
Tabela 11 – Parâmetros utilizados na RBF aplicada na Etapa 2.....	84
Tabela 12 – Parâmetros utilizados na MLP aplicada na Etapa 3.....	87
Tabela 13 – Parâmetros utilizados na RBF aplicada na Etapa 3.....	88

LISTA DE QUADROS

Quadro 1 – Principais conceitos do absentéismo.....	10
Quadro 2 – Principais conceitos do presenteísmo.....	15
Quadro 3 – Relação entre autor (es), palavras-chave, técnicas aplicadas, idioma e base de dados dos artigos selecionados.....	54
Quadro 4 – Softwares utilizados no trabalho.....	57
Quadro 5 – Descrição e origem dos atributos por base de dados.....	60
Quadro 6 – Características dos atributos das bases de dados.....	63
Quadro 7 – Descrição dos atributos utilizados na rede SOM, MLP 1 e MLP 2.....	65
Quadro 8 – Descrição dos atributos para auxílio à previsão do comportamento no trabalho..	68
Quadro 9 – Parâmetros utilizados nas técnicas.....	70
Quadro 10 – Atributos aplicados na rede SOM (14 atributos).....	75
Quadro 11 – Síntese dos experimentos da Etapa 1.....	82
Quadro 12 – Síntese dos experimentos da Etapa 2.....	86
Quadro 13 – Comparação dos experimentos computacionais das Etapas 1 e 2.....	86
Quadro 14 – Síntese dos experimentos computacionais da Etapa 3.....	90
Quadro 15 – Comparação dos experimentos computacionais das Etapas 1, 2 e 3.....	91
Quadro 16 – Comparação dos experimentos computacionais da Etapa 1.....	92
Quadro 17 – Comparação dos experimentos computacionais da Etapa 2.....	94
Quadro 18 – Comparação dos experimentos computacionais da Etapa 3.....	96
Quadro 19 – Consolidação dos experimentos computacionais das Etapa 1, 2 e 3.....	99

LISTA DE ABREVIATURAS E SIGLAS

- ABERGO - Associação Brasileira de *Ergonomia*
- ACO - *Ant Colony Optimization*
- AET - Análise Ergonômica do Trabalho
- BMU - *Best Match Unit*
- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CID - Classificação Internacional de Doenças
- CLT - Consolidação das Leis do Trabalho
- CNAM - *Conservatoire National des Arts et Métiers*
- CPU - *Central Processing Unit*
- DBSCAN - *Density Based Spatial Clustering of Applications with Noise*
- DM - *Data Mining*
- EAA_QVT - Ergonomia da Atividade Aplicada à Qualidade de Vida no Trabalho
- EQ - Erro de Quantização
- ET – Erro Topográfico
- GB - *Gigabyte*
- IEA - *International Ergonomics Association*
- IK – Índice Kappa
- IMC - Índice de massa corporal
- INPI - Instituto Nacional da Propriedade Industrial
- INSS - Instituto Nacional do Seguro Social
- MBP - *Multiple Back-Propagation*
- MLP - *MultiLayer Perceptron*
- OMS - Organização Mundial de Saúde
- QVT - Qualidade de Vida no Trabalho
- RAM - *Random Access Memory*
- RBF - *Radial Basis Function*
- REBA - *Rapid Entire Body Assessment*
- RF – *Random Forest*
- RNA - Rede Neural Artificial
- RS - *Rough Sets*
- SOM - *Self-Organizing Maps*
- TB – *Terabyte*

UCI - University of California Irvine

UFLA - Universidade Federal de Lavras

WEKA - *Waikato Environment for Knowledge Analysis*

WHO - *World Health Organization*

1. INTRODUÇÃO

O homem e o trabalho se interligam de diferentes formas de acordo com cenários políticos, econômicos, sociais, tecnológicos, demográficos e culturais, o que pode ocasionar profundas transformações nas organizações.

Essas transformações geram a necessidade de atender e atingir metas mais exigentes no trabalho, o que aumenta, consideravelmente a sobrecarga física e mental do funcionário. No entanto, essa sobrecarga pode passar despercebida pelo empregador e causar tensões que desencadeiam reações adversas ao interferir diretamente na qualidade de vida no trabalho, levando aos fenômenos chamados de absenteísmo ou presenteísmo (BERNSTROM, 2013; LOHAUS; HABERMANN, 2019).

O absenteísmo é um fenômeno definido como o não comparecimento do empregado ao trabalho de forma habitual e com frequência regular; conseqüentemente, isso causa o descumprimento de suas obrigações (JOHNS, 2008).

O absenteísmo difere da falta ao trabalho, em situação pontual e isolada, em função de sua frequência e habitualidade (JOHNS, 2008; PRATER; SMITH, 2011; LOHAUS; HABERMANN, 2019).

Em paralelo ao absenteísmo, existe outro fenômeno relacionado ao não cumprimento das obrigações, denominado presenteísmo. Neste caso, há a presença do empregado no local de trabalho, mas a realização de suas atividades e de suas funções ocorre de modo improdutivo por diversos motivos, dentre eles doenças físicas ou mentais (BÖCKERMAN; LAUKKANEN, 2010; JOHNS, 2010).

Assim, para identificar o motivo do baixo rendimento dos empregados, faz-se necessária uma análise das causas e conseqüências que podem estar relacionadas a diversos tipos de absenteísmo, bem como de estudos que visam ampliar a percepção do presenteísmo e considerar as situações que alteram a condição de saúde e bem-estar dos empregados (INOUE et al., 2008; BRATSBERG; FEVANG; ROED, 2013; UMANN; GUIDO; SILVA 2014; SHIMABUKU; MENDONÇA; FIDELIS, 2017).

Entre os estudos existentes sobre as causas e conseqüências relacionadas aos tipos de absenteísmo, encontra-se o de Zamanian et al. (2014), que trata das atividades normalmente desempenhadas nas empresas, como movimentos repetitivos, elevação dos braços, puxar, empurrar, transportar, além de outras que são propícias ao desenvolvimento de distúrbios musculoesqueléticos. Em razão disso, a não observação das técnicas de levantamento e abaixamento de peso pode ocasionar lesões e afastamentos por motivos diversos.

Já em relação aos estudos sobre as causas do presenteísmo, encontram-se os de Hemp (2004), Johns (2010) e Umann, Guido e Silva (2014), cujas pesquisas defendem que as causas do presenteísmo nem sempre são aparentes e fáceis de identificar, o que dificulta a percepção dos gestores e, conseqüentemente, a tomada de medidas necessárias para prevenir ou minimizar os seus efeitos. Tais estudos apontam para problemas de saúde, organizacionais ou pessoais, não ligados à saúde, bem como para custos, uma vez que impacta na produtividade da empresa. Por esta e outras razões, o presenteísmo vem sendo cada vez mais estudado.

O absenteísmo ou presenteísmo, causado por distúrbios osteomusculares, tornou-se um grande problema para as empresas (ZAVARIZZI; ALENCAR, 2018), haja vista que as atividades laborais são diversificadas e que cada profissional adota uma determinada postura. Varrer o chão, por exemplo, exige do corpo a ação conjunta de diversos ossos, músculos e articulações que, ao serem utilizados de forma inadequada, podem causar lesão. Sendo assim, nota-se a necessidade de um estudo sobre as causas e conseqüências que podem estar relacionadas ao absenteísmo (BRATSBERG; FEVANG; ROED, 2013; IIDA; BUARQUE, 2016).

Nesse cenário, aspectos antropométricos e ergonômicos são importantes dentro do contexto do absenteísmo, principalmente em atividades classificadas como trabalho pesado e com um alto índice de repetitividade. A antropometria trata das medidas físicas, ou seja, tamanho e proporções do corpo humano. A população é composta de indivíduos de diferentes tipos físicos que apresentam distinções nas proporções de cada segmento do corpo. De modo geral, a antropometria trata de medidas físicas corporais e verifica seu grau de adequação aos postos de trabalho (KROEMER; GRANDJEAN, 2005; FONSECA; SANTOS, 2013; IIDA; BUARQUE, 2016).

A ergonomia, por sua vez, estuda a adaptação do trabalho ao homem e o comportamento humano no trabalho, sendo necessário garantir boas condições laborais, a fim de manter e promover a saúde no ambiente corporativo. Sua evolução está relacionada às transformações socioeconômicas e, sobretudo, tecnológicas (KROEMER; GRANDJEAN, 2005).

A ergonomia veio integrar o conhecimento proveniente das ciências humanas para adaptar tarefas, sistemas, produtos e ambientes às habilidades e limitações físicas e mentais das pessoas. Em outras palavras, consiste em promover a melhoria do trabalho e das condições de vida dos empregados, desenvolvendo conhecimento e apoio a análises e projetos de produtos e de situações produtivas (KARWOWSKI, 2006; SANTOS et al., 2016).

A detecção e a prevenção do absenteísmo e do presenteísmo são importantes, uma vez que passam pela análise das informações geradas pelos empregados, e devem ser registradas de

forma completa e abrangente como, por exemplo: registros de Classificação Internacional de Doenças (CID), quando realizados exames médicos e fisioterapia (BUSTILLOS; VARGAS III; GOMERO-CUADRA, 2015). Fazer o registro das faltas ao trabalho, dos motivos alegados ou das justificativas para essas faltas, bem como fazer o registro dos atrasos ou saídas antecipadas protagonizadas pelos trabalhadores de uma organização, pode gerar dados importantes ao processo de análise e diagnóstico organizacional do absenteísmo; o acompanhamento e o monitoramento são um importante elemento de diagnóstico e facilitador do desenvolvimento de intervenções nas organizações (FLORES et al., 2016).

Prever tendências tem sido desafiador para as empresas, já que são muitas as informações armazenadas em bases de dados. Nesses grandes volumes de dados, escondem-se diversas relações interessantes que, ao serem descobertas, ampliam conhecimento para apoiar a tomada de decisão.

Assim, a mineração de dados, uma área da Inteligência Artificial, é responsável pela extração e análise do conhecimento implícito e útil contido em uma base de dados, possibilitando descobrir padrões, regras e informações que permitam montar estratégias corporativas eficientes utilizando, para isso, uma técnica da Inteligência Artificial (IA) como a rede neural artificial (RNA) (LINOFF; BERRY, 2011; BRAHAMI; ATMANI; MATTA, 2013).

Diversas técnicas, como as Redes Neurais Artificiais (RNAs), são utilizadas na mineração de dados, pois tratam-se de modelos constituídos por unidades de processamento simples, chamados de neurônios artificiais, que calculam funções matemáticas. Esses modelos são inspirados na estrutura do cérebro e têm como objetivo simular o comportamento humano com capacidade de aprendizado, generalização, associação e abstração; quando submetidas a treinamento, as redes são particularmente eficientes para o mapeamento (entrada/saída) de sistemas não lineares, constituído por sistemas paralelos distribuídos em compostos de unidades simples de processamento, além de simular sistemas complexos (HAYKIN, 2001).

Dentre as várias arquiteturas disponíveis de RNAs, existem três que são utilizadas em diversas aplicações: a *Multilayer Perceptron* (MLP) (NAIK; SAMANT, 2016), a *Radial Basis Function* (RBF) (HAN; PEI; KAMBER, 2012) e a *Self-Organizing Map* (SOM) (KOHONEN, 2013).

Outras técnicas aplicadas em mineração de dados são: *Random Forest* (RF), ou Florestas Aleatórias (BREIMAN; CUTLER, 2014) e *Ant Colony Optimization* (ACO), ou Algoritmo de Otimização por Colônia de Formigas.

A RF é um algoritmo de classificação e regressão não paramétrico e não linear, que gera um conjunto de árvores de decisão. Nos problemas de classificação, o resultado obtido é dado

por um sistema de votação, em que a nova classe predita é aquela que obteve o maior número de votos; ou seja, segue o conceito de aprendizado por conjunto (BREIMAN, 2001).

O ACO é uma metaheurística inspirada no comportamento das colônias de formigas na natureza (DORIGO; STÜTZLE, 2004). O algoritmo ACO para classificação cria regras de classificação SE-ENTÃO. A parte SE corresponde ao antecedente da regra e contém termos que representam testes nos valores dos atributos. A parte ENTÃO corresponde ao consequente da regra e contém uma previsão de valor da classe (PARPINELLI; LOPES; FREITAS, 2002).

O desenvolvimento e a aplicação de uma pesquisa apoiada nas técnicas de mineração de dados, para previsão dos comportamentos presenteísta, normal e absenteísta, mostram-se relevantes para a ampliação dos estudos sobre o assunto. Além disso, a aplicação dessas técnicas, sobre os dados ergonômicos e antropométricos, é relevante quando associadas aos dados relativos ao absenteísmo e presenteísmo.

1.1 IDENTIFICAÇÃO DE LACUNA DE PESQUISA

Segundo Gaudart (2016), são altos os índices de aposentadorias por invalidez ou afastamentos por problemas causados pela realização de atividades laborais de maneira errônea. De acordo com a Secretaria de Previdência (2020), em 2017, a doença que mais afastou os brasileiros dos postos de trabalho foi a dorsalgia (dor nas costas ou doença da coluna). Foram diagnosticados aproximadamente 83,8 mil casos, sendo a enfermidade mais frequente nos últimos anos e a principal responsável pelos auxílios-doença concedidos pelo Instituto Nacional do Seguro Social (INSS), seguida pelas fraturas nas pernas e no punho.

Tais problemas causados por determinadas atividades, além de outras consequências, podem afetar o desempenho da função social, como perda do condicionamento físico, deterioração da saúde geral, diminuição da participação em atividades sociais, estresse familiar, diminuição do contato com a comunidade, irritação, ansiedade e depressão (FERREIRA; NAVEGA, 2010).

A pesquisa bibliográfica realizada neste estudo permitiu observar que o absenteísmo é um dos assuntos mais estudados no comportamento organizacional, devido em grande parte às mudanças no ambiente de trabalho que resulta em empregados sobrecarregados e com menos recursos e suporte organizacional (DA SILVA et al., 2018; MIRAGLIA; JOHNS, 2018).

O presenteísmo ganhou merecida atenção, recentemente como um fator importante que afeta o desempenho organizacional (LOHAUS; HABERMANN, 2019). Embora o assunto esteja ganhando relevância, há lacunas a respeito do entendimento científico sobre esse campo

de estudo em razão da escassez de produções nacionais e internacionais que apontem suas causas e seus efeitos (SHIMABUKU; MENDONÇA; FIDELIS, 2017).

Lohaus e Habermann (2019) alertam que tanto o absenteísmo quanto o presenteísmo são vistos pelos pesquisadores como fenômenos relacionados. Além das doenças relatadas por muitos empregados em decorrência do estresse, pode-se observar outros fatores que também levam ao presenteísmo:

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

O comportamento presenteísta pode ser manifestado, segundo níveis variáveis de atuação de fatores pessoais e organizacionais, ou seja, sob influência de fatores pessoais e organizacionais, tem-se um comportamento que, subterraneamente, ameaça a Sustentabilidade Organizacional, isto é, prejudica as melhores relações entre a organização e seus empregados (SILVEIRA, et al., 2015; GARRIDO et al, 2017).

Esses fatores diferentes que levam ao presenteísmo podem desencadear doenças que levam ao absenteísmo.

As CIDs (M00-M99), que se refere a doenças do sistema osteomuscular e do tecido conjuntivo, e (S00-T98), que se refere a lesões, envenenamento e algumas outras consequências de causas externas, são um atributo de relacionamento entre a antropometria, ergonomia, absenteísmo e o presenteísmo, sendo todas relacionadas a trabalhos que envolvem atividades com manuseio de cargas: levantar, abaixar, empurrar, puxar, carregar, segurar e arrastar; em geral, são atividades que costumam exigir bastante esforço físico.

Investigar as características ergonômicas e antropométricas que mais se ajustam ao tipo de trabalho é um assunto que vem sendo bastante discutido na literatura (DUL; WEERDMEESTER, 2004; KROEMER; GRANDJEAN, 2005; DIONNE et al., 2008; OLIVEIRA et al., 2015; MEZIAT FILHO; SILVA, 2011; IIDA; BUARQUE, 2016).

O levantamento bibliográfico aqui realizado, que pode ser consultado no capítulo 3 e na seção 3.2, permitiu observar que os estudos sobre o absenteísmo e o presenteísmo, envolvendo ergonomia e antropometria, são pouco abordados. Tal fato comprova a contribuição teórica e prática deste estudo para a academia e as organizações, uma vez que se trata de uma das pesquisas pioneiras em associar os quatro assuntos.

Somando ainda mais ao estudo, tem-se a aplicação da mineração de dados, que reforça a lacuna de pesquisa, já que, no levantamento bibliográfico realizado, não foi encontrado nenhum trabalho sobre absenteísmo e presenteísmo, com foco em ergonomia e antropometria, utilizando qualquer técnica de mineração de dados, ciência de dados ou inteligência artificial.

A importância deste estudo está na sua contribuição para a saúde dos empregados, para a verificação de equipamentos adequados, para a implantação de programas voltados para a qualidade de vida no trabalho e para o acompanhamento do empregado durante suas atividades, com vista à previsão do seu comportamento. Para as empresas, a contribuição está na possibilidade de manter os níveis de absenteísmo e presenteísmo baixos, de modo a promover mais qualidade e produtividade nos serviços.

1.2 JUSTIFICATIVA E MOTIVAÇÃO

A decisão de prever os comportamentos presenteísta, normal e absenteísta foi motivada pela necessidade de auxiliar as empresas a manterem os níveis de absenteísmo e presenteísmo mais baixos, uma vez que o baixo desempenho da mão de obra reflete diretamente na competitividade e lucratividade da empresa. Os reflexos do absenteísmo e presenteísmo são extremamente prejudiciais para a organização (LOPES et al., 2017).

Em relação aos empregados, pode-se elencar: a sobrecarga de trabalho, o excesso de horas extras, e a percepção equivocada (ou não) de falta de atuação da liderança, o que pode causar revolta no grupo produtivo, o adoecimento de outros empregados e a desmotivação para o trabalho.

Já no que se refere ao ambiente de trabalho, pode-se elencar: o prejuízo ao clima organizacional da empresa e o prejuízo à gestão. Para o negócio da empresa, pode-se elencar: a diminuição da qualidade do serviço ou produto, a redução da capacidade de atender o mercado cada vez mais exigente, a redução da produtividade e competitividade e, por conseguinte, o aumento de custos, gerando assim serviços e produtos mais caros e com menor qualidade, o que pode ocasionar perda de clientes.

Santos et al. (2016) argumentam que, na perspectiva da ergonomia situada, considerar a atividade significa não levar em conta as funções gerais, como posturas e movimentos, de forma isolada, mas sim os comportamentos e gestos, tais como se apresentam naturalmente nas situações particulares de trabalho, existentes ou em concepção.

A pesquisa sobre absenteísmo, presenteísmo, antropometria e ergonomia também foi motivada pela sua importância para a saúde dos empregados, para a verificação de

equipamentos adequados, para a implantação de programas voltados para a qualidade de vida no trabalho, e para a qualidade e produtividade dos serviços.

Neste sentido, dados antropométricos e ergonômicos são importantes quando relacionados ao absenteísmo e ao presenteísmo, principalmente em atividades classificadas como trabalho pesado e com um alto índice de atividades repetitivas (IIDA; BUARQUE, 2016).

Esta pesquisa também teve como motivação a lacuna existente sobre o assunto, que se mostrou pouco explorado nas pesquisas, conforme pôde-se constatar com o levantamento bibliográfico realizado neste trabalho. A escassez do tema permitiu a aplicação de mineração de dados para prever os comportamentos dos empregados presenteístas, normais e absenteístas. Além disso, o conhecimento extraído pela mineração de dados ajudou a transformar os resultados da pesquisa em ações concretas para melhorar a qualidade de vida e o trabalho dos empregados.

Vale destacar que, este tipo de aplicação vem ao encontro à motivação que ocorre, em termos mundiais, a respeito de usar técnicas inteligentes na resolução de problemas de diversas áreas. Isto reflete de certa forma a atualidade e importância do trabalho.

1.3 PROBLEMA DE PESQUISA

Os excessivos custos com assistência médica, tratamentos, afastamentos e reintegração ao trabalho, que podem refletir em altos índices de absenteísmo, são o principal motivo de preocupação dos gestores das empresas, que estão em busca de soluções efetivas para resolver esses problemas de forte influência social e financeira.

No Brasil, o comprometimento com a produção, que na maioria das vezes está voltada para a produtividade, resulta na intensificação do trabalho, gerando condições desfavoráveis ao trabalhador. Além do mais, a Norma Regulamentadora 17 (NR17), que trata de forma específica das questões ergonômicas no ambiente de trabalho exige das empresas a análise ergonômica para investigar se as condições laborais estão adequadas às características psicofisiológicas dos trabalhadores, nos aspectos levantamento, transporte e descarga de materiais, mobiliário, equipamentos, condições ambientais do posto de trabalho e organização de trabalho (PINTO; TERESO; ABRAHÃO, 2018).

Outra preocupação das empresas é com o presenteísmo. Estudos empíricos identificaram correlatos desse fenômeno, no entanto, uma clara distinção entre suas causas e seus efeitos ainda precisa ser feita. Lohaus e Habermann (2019) mostram em suas pesquisas que, até hoje, nem uma definição uniforme nem métodos consistentes de mensuração foram

empregados na pesquisa sobre presenteísmo. Embora tenha havido várias tentativas de desenvolvimento de teoria, não foi estabelecido nenhum modelo abrangente que pode explicar por que as pessoas escolhem presenteísmo ou absenteísmo.

Até o momento, a análise de Miraglia e Johns (2016) é a revisão mais abrangente dos correlatos do presenteísmo. No entanto, mais pesquisas se fazem necessárias para separar as causas dos efeitos do presenteísmo. Além disso, em contraste com o absenteísmo, o presenteísmo ainda não foi formalmente registrado (HANSEN; ANDERSEN, 2008).

Segundo Evans-Lacko e Knapp (2016), é mais difícil de se determinar que o custo do presenteísmo excede o custo do absenteísmo, fato este que serve de inspiração para mais pesquisas sobre métodos consistentes de previsão. De acordo com Johns (2010), o fenômeno presenteísmo é importante demais, por razões teóricas e práticas, para ser deixado apenas nas mãos de pesquisadores médicos e consultores de saúde.

Assim, a **questão de pesquisa** desta tese é a seguinte: “A mineração de dados aplicada em base de dados antropométricos, ergonômicos, absenteístas e presenteístas pode auxiliar na previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no trabalho?”

1.4 OBJETIVOS

1.4.1 Objetivo geral

O objetivo geral deste estudo foi investigar a aplicação da mineração de dados em base de dados antropométricos, ergonômicos, absenteístas e presenteístas para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no trabalho.

1.4.2 Objetivos específicos

- Selecionar e enriquecer as bases de dados oriundas da literatura para a realização dos experimentos computacionais;
- Aplicar as técnicas de mineração de dados selecionadas: *Multilayer Perceptron* (MLP), *Radial Basis Function* (RBF), *Self-Organizing Maps* (SOM), *Random Forest* (RF) e *Ant Colony Optimization* (ACO);
- Comparar os resultados computacionais obtidos com a aplicação das técnicas selecionadas para prever o comportamento dos empregados.

1.5 DELIMITAÇÃO DO TEMA DE PESQUISA

Para compor o estudo, selecionaram-se três bases de dados oriundas de uma tese de doutorado, uma dissertação de mestrado e um artigo científico, que juntas continham os atributos para a realização dos experimentos computacionais.

A base de dados oriunda da tese de doutorado, desenvolvida por Ferreira (2019), foi escolhida por tratar-se de uma base composta por atributos relacionados ao absenteísmo e presenteísmo, além de possuir atributos relacionados à antropometria. Trata-se da única base de dados de acesso livre que contempla dados de empregados brasileiros.

A base de dados oriunda da dissertação de mestrado, realizada por Defani (2007), foi escolhida por tratar-se de uma base composta por atributos relacionados a medidas antropométricas de empregados brasileiros.

Já a base de dados oriunda do artigo científico, desenvolvido por Martiniano, Ferreira e Sassi (2018), foi escolhida por tratar-se de uma base composta por atributos relacionados à ergonomia e por pertencer à mesma empresa estudada por Ferreira (2019); ou seja, os empregados executam as mesmas atividades laborais e passaram pelos mesmos treinamentos corporativos relacionados à postura ergonômica.

A seleção das técnicas aplicadas baseou-se na vasta gama de aplicações das redes neurais artificiais em mineração de dados. A RF e o ACO foram selecionados para serem aplicados em problemas de previsão. Essas técnicas, por serem distintas e desenvolvidas recentemente, foram aplicadas para evidenciar as diferenças no desempenho e no tempo de processamento de cada uma delas.

1.6 ORGANIZAÇÃO DO TRABALHO

Este estudo foi estruturado em cinco capítulos. A Introdução é apresentada no capítulo 1. No capítulo 2, apresenta-se a fundamentação teórica sobre absenteísmo, presenteísmo, antropometria, ergonomia e mineração de dados. Já no capítulo 3, são abordados os materiais e métodos para desenvolvimento deste estudo. No capítulo 4, apresentam-se os resultados dos experimentos computacionais e, por fim, no capítulo 5, a conclusão.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se a fundamentação teórica sobre absenteísmo, presenteísmo, antropometria, ergonomia, além da mineração de dados e suas técnicas.

2.1 ABSENTEÍSMO

O absenteísmo difere da falta ao trabalho porque esta é uma situação pontual e isolada em função de sua frequência e habitualidade (JOHNS, 2008; PRATER; SMITH, 2011; LOHAUS; HABERMANN, 2019).

O absenteísmo é conceituado de diversas formas. No Quadro 1, pode-se verificar as diferentes conceituações realizadas pelos principais autores no assunto.

Quadro 1 – Principais conceitos do absenteísmo

Autores	ABSENTEÍSMO
Silva e Marziale (2003)	Qualquer ausência do empregado, até mesmo os afastamentos legais, que são de direito do empregado, tais como licença maternidade, férias, abonos, etc.
Johns (2008)	Fenômeno definido como o não comparecimento do empregado ao trabalho de forma habitual e com frequência regular; conseqüentemente, é o não cumprimento das obrigações.
Johns (2010)	Conceituado como um fenômeno, o absenteísmo é relacionado ao não comparecimento do empregado ao trabalho de forma frequente e habitual, o que geralmente indica uma diminuição considerável do desempenho funcional no local de trabalho ou que o serviço não está sendo realizado.
Lohaus e Habermann (2019)	Fenômeno de não comparecer ao trabalho com frequência. Tem sido amplamente pesquisado nas últimas décadas devido ao impacto na produtividade das empresas.

Fonte: Elaborado pelo autor (2020).

Além dos vários conceitos, que reflete a multidisciplinaridade de conhecimentos disponíveis sobre o assunto, o absenteísmo consiste em uma resposta comportamental de interesse, uma vez que esses conhecimentos normalmente não se complementam isoladamente, sendo desenvolvidos à parte em cada uma das disciplinas, que vão da Psicologia à Engenharia, passando por diversas áreas (JOHNS 2003).

Quick e Lapertosa (1982) e Inoue (2008) classificam o absenteísmo em cinco categorias: 1) absenteísmo por doença: justificada por atestado ou licença-saúde; 2) absenteísmo por patologia profissional: causado por acidente de trabalho e/ou doença profissional; 3) absenteísmo legal: respaldado pela Consolidação das Leis do Trabalho (CLT), como gestação,

nojo (luto), gala (casamento), doação de sangue e serviço militar; 4) absenteísmo compulsório: por suspensão disciplinar, por prisão ou por outro impedimento de comparecer ao trabalho; e 5) absenteísmo voluntário: por razões particulares das quais não são justificadas por doença.

2.1.1 O absenteísmo nas empresas

O interesse na causa do absenteísmo aumentou nas últimas três décadas devido ao alto custo para a empresa e à necessidade de melhoria da qualidade de vida no trabalho por parte dos empregados, para acomodar as novas necessidades do mercado e dos empregados. Em resposta às tendências da globalização, as organizações vêm introduzindo mudanças substanciais no design do trabalho, como o horário flexível e o trabalho remoto. Isso certamente afeta os relacionamentos no local de trabalho (MIRAGLIA; JOHNS, 2018).

Reduzir ao máximo suas perdas tem sido um dos maiores desafios, por exemplo, para o setor de saúde ocupacional, dada a dimensão do fenômeno que engloba a detecção e o monitoramento das causas de fundo psicológico, físico e/ou ambiental, considerando as circunstâncias que levam o empregado ao absenteísmo (ADDAE; JOHNS; BOIES, 2013; RUSSO et al., 2013).

O absenteísmo muito elevado em uma empresa gera perdas de produtividade, dificulta a gestão do trabalho e, por conseguinte, cria grupos desmotivados (HALBESLEBEN; WHITMAN; CRAWFORD, 2014).

Certas organizações consideram como absenteísmo a falta injustificada e imprevista do empregado, já outras, pensam no absenteísmo como qualquer ausência do empregado (SILVA; MARZIALE, 2003). De qualquer forma, cada organização considera o absenteísmo de acordo com sua realidade e cultura organizacional, bem como a realidade de seus empregados.

De acordo com Garrido et al. (2019) acentuado interesse tem sido direcionado ao presenteísmo, devido ao impacto para o desempenho econômico e operacional das organizações, para os empregados e para a sociedade. A busca por abordagens que permitam diagnosticar e quantificar os impactos do presenteísmo tem estimulado a investigação dos múltiplos fatores que o influenciam e são influenciados por ele e que afetam a produtividade.

Assim, Garrido et al. (2019) propuseram, diagnosticar a realidade da organização em termos de presenteísmo e cooperação por meio da aplicação de métodos analíticos de uma organização pública de pesquisa. Os critérios metodológicos para mensuração envolveram o uso de métricas e técnicas analíticas que permitiram diagnosticar os seguintes resultados: 1 - níveis gerais de presenteísmo e cooperação; 2 - níveis de presenteísmo e cooperação por equipe

estudada; 3 - correlações entre o presenteísmo e a cooperação; e 4 - possíveis fatores de influência nos níveis de presenteísmo e cooperação. O estudo fundamentou e mostrou uma evidência empírica da relação entre o presenteísmo e a cooperação e validou meios para diagnosticar e quantificar seus impactos na realidade organizacional.

Contudo, para Britt e Jex (2015), conciliar a expectativa do empregado com as necessidades organizacionais é um desafio diante das dificuldades apresentadas pelo ambiente de trabalho, como as situações de insalubridade, longas jornadas de trabalho, baixa remuneração, duplo emprego, tensão emocional e estresse, que podem ocasionar o absenteísmo.

Mesmo para reduzir o absenteísmo, vale lembrar os aspectos voltados à Qualidade de Vida no Trabalho (QVT), que estão ligados à produtividade da organização. Mudanças no estilo de vida, equilíbrio entre trabalho e vida pessoal, saúde e condicionamento físico, antes não considerados benefícios, são vistos agora como práticas primordiais para os empregados. Sendo assim, a gestão, que dita como maximizar a produtividade dos empregados, concentra-se em duas áreas principais: motivação pessoal e infraestrutura do ambiente de trabalho (CHANDRASEKAR, 2011).

Para Klein, Pereira e Lemos (2019), melhorar a qualidade de vida de seus empregados torna a organização mais saudável, competitiva e produtiva, e essa é a principal função da QVT, que está focalizada no potencial humano e no meio que convive, em todos os sentidos. Por meio de habilidades, técnicas, conhecimentos e trabalho, geram o bem ou serviço que permite a elas atingir os objetivos perante a sociedade e sobreviver. As empresas que buscam o sucesso de mercado, passam a ter uma preocupação maior com relação aos seus empregados e com o ambiente de trabalho. O desenvolvimento das organizações depende das pessoas envolvidas para o alcance da missão, das metas e dos resultados.

Baseando-se nisso, estudos voltados para a preservação da integridade física e mental dos empregados devem ser incentivados e estar em sintonia com uma concepção compartilhada e negociada de gestão do trabalho, uma vez que podem contribuir com ações de prevenção à saúde e de redução do absenteísmo, pois quanto mais motivados e satisfeitos, mais os empregados tendem a produzir (CHANDRASEKAR, 2011; SILVA, FURTADO e ZANINI, 2015).

2.1.2 Pesquisas relacionadas ao absenteísmo

O interesse nas pesquisas sobre a causa do absenteísmo se deu, em parte, à grande diversidade de métodos de pesquisa que foram empregados para estudar o fenômeno. Os métodos melhoraram a compreensão sobre o absenteísmo no trabalho. Além disso, considera-

se que quase todas as pessoas passaram pelo absenteísmo em algum momento ao longo de suas vidas profissionais, e que a diversidade metodológica se expandiu para uma ampla variedade de disciplinas acadêmicas e profissionais que se interessaram pelo absenteísmo (PATTON; JOHNS, 2012).

Os motivos das pesquisas são variados. Addae, Johns e Boies (2013); Da Silva et al., (2018) e Lohaus e Habermann (2019) trataram do absenteísmo por saúde e doenças ocupacionais; Trounson e Pfeifer (2017), do bem-estar psicológico; Assunção e Abreu (2017), Haeffner et al. (2018), Zavarizzi e Alencar (2018) trataram do absenteísmo por distúrbio musculoesquelético; Ferro et al., (2018), do absenteísmo por condição de Trabalho em equipe, e Miraglia e Johns (2018) do absenteísmo relacionado ao contexto social.

Flores et al. (2016) destacam o absenteísmo e suas consequências tanto para o empregado como para a organização. As principais causas permanecem pobremente documentadas, o que dificulta a elaboração de programas preventivos e de reabilitação dos empregados, principalmente nos países em desenvolvimento.

Flores et al. (2016) também destacam em suas pesquisas que não há espaços para práticas de gestão isoladas, descontextualizadas e desalinhadas com os propósitos das organizações. Por isso, elas devem voltar-se não só para resultados econômicos, mas, sobretudo, para a saúde e o bem-estar dos empregados. Além disso, é prática imprescindível, dentro das organizações, detectar e compreender o que favorece o absenteísmo.

Lohaus e Habermann (2019) ressaltam o absenteísmo como o fenômeno de não comparecer ao trabalho com frequência; tema este que tem sido amplamente pesquisado nas últimas décadas devido ao seu impacto na produtividade das empresas.

Dentre os estudos existentes sobre o absenteísmo, destaca-se um modelo multinível da dinâmica social do fenômeno do absenteísmo criado por Miraglia e Johns (2018), que classificam as forças do contexto social e moldam o comportamento dos empregados em sete níveis: indivíduo, grupo de trabalho, referência de vida, organização, ocupação, comunidade/localidade e nação/sociedade. O modelo facilita a compreensão sobre quais são os atores sociais dentro e fora da organização, que influenciam os padrões e comportamentos de absenteísmo individuais e como essa influência social é exercida.

Shuang et al. (2019) desenvolveram um estudo sobre o estresse térmico, um fator que afeta a saúde dos trabalhadores, causa absenteísmo e reduz a capacidade de trabalho, diminuindo assim a produtividade. O estresse térmico pode ter efeitos econômicos, já que a saúde da força de trabalho é o insumo primário para a produção econômica. Além disso, o efeito do estresse térmico na produtividade do trabalho também é distribuído de forma desigual devido

à diferente adaptabilidade do calor (SHUANG et al., 2019).

Os resultados mostraram que sob os atuais alvos globais de aquecimento médio, o estresse térmico afetará a maioria das regiões do mundo. As perdas de produtividade do trabalho relacionadas ao calor, para as metas de aquecimento global de 1,5 °C, 2 °C, 3 °C e 4 °C graus celsius (°C), nos cenários apresentados, também mostraram um padrão global, que pode ajudar a quantificar o quanto os países em desenvolvimento serão mais afetados pelo aquecimento do que os países desenvolvidos (SHUANG et al., 2019).

Drakopoulos e Grimani (2013) sugerem em suas pesquisas que o absenteísmo e a satisfação no trabalho podem estar mais fortemente relacionados às condições de trabalho. No entanto, falta na literatura atenção para absenteísmo por lesões. Esse artigo tenta preencher essa lacuna, e examina o efeito da satisfação no trabalho sobre o absenteísmo relativo a lesões, usando dados gregos e britânicos.

Os resultados mostraram que em termos de possíveis implicações políticas, os fatores determinantes cruciais da satisfação no trabalho, como salário, autonomia, oportunidades de promoção, qualidade da supervisão e boas condições de trabalho, que têm um alto impacto no aumento da satisfação no trabalho, também podem levar a uma redução do absenteísmo por lesões, além de menores custos de absenteísmo e aumento da produtividade. No entanto, mais pesquisas e comparações entre países sobre as semelhanças e diferenças do absenteísmo são claramente necessárias para uma melhor compreensão da questão (DRAKOPOULOS; GRIMANI, 2013).

Trounson, Pfeifer e Skues (2018) propuseram um estudo que examina como as respostas de 174 agentes penitenciários à adversidade moderam a associação entre a adversidade percebida no local de trabalho e o bem-estar psicológico, com o objetivo de avaliar o grau em que a adversidade percebida no local de trabalho prevê o bem-estar psicológico do agente penitenciário e o impacto subsequente em resultados organizacionais negativos, como absenteísmo, presenteísmo e insatisfação no trabalho. O estudo mostrou a relação entre a adversidade percebida no local de trabalho, o bem-estar psicológico do empregado e a ocorrência de impactos organizacionais negativos, como absenteísmo, presenteísmo e insatisfação no trabalho.

Trounson, Pfeifer e Skues (2018) concluíram em seu trabalho que a associação identificada entre a percepção de adversidade no local de trabalho e o bem-estar psicológico fornece evidências para sugerir que os agentes penitenciários que percebem um nível elevado de adversidade em seu ambiente de trabalho têm maior probabilidade de relatar uma menor sensação de bem-estar psicológico e são menos propensos a relatar uma sensação de

prosperidade e são mais propensos a sentir angústia.

2.2 PRESENTEÍSMO

Outro fenômeno que merece ser considerado como um fator importante que afeta o desempenho organizacional chama-se presenteísmo.

O presenteísmo é conceituado de diversas formas. No Quadro 2, pode-se verificar as diferentes conceituações realizadas pelos principais autores no assunto.

Quadro 2 – Principais conceitos do presenteísmo

Autores	PRESENTEÍSMO
Hemp (2004); Johns (2006); Böckerman e Laukkanen (2009, 2010); Lohaus e Habermann (2019)	Fato de os empregados estarem no local de trabalho, mas, devido a problemas de ordem física ou psicológica, não cumprirem por completo as suas funções.
Franco, Druck e Seligmann-Silva (2010)	Ausência mental ou emocional no trabalho, decorrente de problemas de saúde, financeiros, familiares e outros, que não exigem o afastamento do empregado, mas reduzem seu rendimento.
Martinez e Ferreira (2011)	Configura-se como um ato de cidadania institucional quando os empregados acreditam que as organizações valorizam sua presença.
Shimabuku, Mendonça e Fidelis (2017)	O presenteísmo foi considerado algum tempo como o absenteísmo de corpo presente, muito provavelmente pelo fato de que o absenteísmo e o presenteísmo tendem a possuir uma base comum no que se refere aos fatores psicossociais.

Fonte: Elaborado pelo autor (2020).

Em contraste com o absenteísmo, o presenteísmo apenas recentemente ganhou atenção como um fator importante que afeta o desempenho organizacional (UMANN; GUIDO; SIIVA 2014; LOHAUS; HABERMANN, 2019). Embora o presenteísmo esteja ganhando relevância, alguns fundamentos importantes sobre o assunto ainda precisam ser esclarecidos satisfatoriamente, em razão da escassez de produções nacionais e internacionais que apontem as causas dos seus efeitos (SHIMABUKU; MENDONÇA; FIDELIS, 2017; LOHAUS; HABERMANN, 2019).

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 (MARTINIANO, 2012).

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, as relações e as condições organizacionais do trabalho, em que 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 (LARANJEIRA, 2009; UMANN; GUIDO; SIIVA, 2014)

Embora consista em um fenômeno cujas ocorrências vão além daquilo que é percebido e mensurável, constata-se que o presenteísmo se porta como um mal ainda mais pernicioso que o absenteísmo (LOPES et al., 2017).

Pesquisas levantadas por Johansen, Aronsson e Marklund (2014) e Zhou et al. (2016) revelam que não se deve presumir que o presenteísta está necessariamente associado à produtividade reduzida, tampouco impedir seus efeitos positivos porque, em alguns casos, o presenteísmo tem resultados estabilizadores de personalidade e efeitos de apoio sobre o status social e econômico do empregado.

Dentre as diversas pesquisas existentes sobre o presenteísmo, há o estudo de Johns (2010), que aponta os fatores positivos de relacionamento entre empregado e equipe, como o sentimento de pertencer ao grupo; essa perspectiva de troca gera sentimentos de recompensa que validam a permanência no local de trabalho, ainda que doente, o que corrobora para o presenteísmo.

No entanto, Wan, Downey e Stough (2014) alertam para outra forma de presenteísmo, ainda que os estudos sobre essa forma de fenômeno sejam limitados. Trata-se do presenteísmo não relacionado ao trabalho, mas que pode impactar financeiramente as organizações, mais até do que o absenteísmo.

O presenteísmo não relacionado ao trabalho refere-se ao comportamento dos empregados que se envolvem em atividades pessoais em vez de atividades relacionadas ao trabalho durante o trabalho (WAN; DOWNEY; STOUGH, 2014).

2.2.1 O presenteísmo nas empresas

Para Umann, Guido e Silva (2014), avaliar o presenteísmo é um desafio, por ser uma condição de difícil mensuração, aparentemente encoberta, que necessita do reconhecimento do empregado sobre sua condição, para o desenvolvimento das atividades no trabalho. Os autores consideram também um desafio a incorporação desse referencial na instância administrativa das instituições, por ter repercussões que, muitas vezes, tornam-se visíveis com a ocorrência de

faltas e afastamentos. Além disso, a medição direta da produtividade é uma tarefa complicada, principalmente para ocupações de trabalho mental e cognitivo, já que engloba fatores de natureza subjetiva.

Para Pichler e Ziebarth (2017), a identificação empírica do comportamento de presenteísmo é extremamente desafiadora, mas crucial para testar uma das principais justificativas econômicas para o pagamento por doença, feito pelo poder público: as externalidades negativas associadas ao presenteísmo contagioso. O presenteísmo contagioso acontece quando empregados com doenças infecciosas vão trabalhar doentes e infectam colegas de trabalho e clientes. Tal comportamento é um problema de saúde pública e uma força motriz da disseminação de doenças contagiosas. Se o contágio é inobservável, o que é geralmente o caso no início dos episódios de doença, a regulação estatal pode obrigar os empregadores a fornecerem incentivos para os empregados ficarem em casa quando doentes para se tratarem; assim, os esquemas públicos de pagamento por doença reduziram o presenteísmo contagioso e a disseminação de doenças para outros empregados.

2.2.2 Pesquisas relacionadas ao presenteísmo

O presenteísmo foi considerado por um tempo como o absenteísmo de corpo presente, muito provavelmente pelo fato de que o absenteísmo e o presenteísmo tendem a possuir uma base comum no que se refere aos fatores psicossociais (SHIMABUKU; MENDONÇA; FIDELIS, 2017).

Estudos em diferentes países e entre distintos grupos ocupacionais mostraram que uma grande parcela de empregados foi trabalhar quando doente. As razões mais relatadas para o presenteísmo foram: não quero sobrecarregar meus colegas; irão desfrutar do meu trabalho e; ninguém é capaz de cumprir minhas responsabilidades. Houve diferenças significativas entre os entrevistados de dois países pesquisados, Noruega e Suécia: uma proporção maior de entrevistados noruegueses aponta para os benefícios de ir para o trabalho apesar da doença, enquanto uma proporção maior de entrevistados suecos relata as consequências econômicas do presenteísmo (JOHANSEN; ARONSSON; MARKLUND, 2014).

Ainda sobre as pesquisas de Johansen, Aronsson e Marklund (2014), os resultados mostraram que quando os entrevistados relatam que praticam presenteísmo porque gostam de seu trabalho, isso geralmente pode ser visto como não problemático. No entanto, vários estudos descobriram que ser presenteísta pode levar a problemas futuros de saúde, e os serviços de saúde ocupacional podem, portanto, considerá-lo como um indicador precoce de produtividade reduzida (JOHANSEN; ARONSSON; MARKLUND, 2014).

Shimabuku, Mendonça e Fidelis (2017), ao traçarem um panorama da atualidade sobre a relação entre os aspectos psicossociais do trabalho e o presenteísmo dos empregados, perceberam uma forte relação entre demandas, controle e apoio social no trabalho e o presenteísmo, constatando assim que as altas demandas de trabalho estão associadas positivamente ao presenteísmo. No entanto, há lacunas a respeito do entendimento científico sobre esse campo de estudo decorrentes da escassez de produções nacionais e internacionais sobre os antecedentes do presenteísmo. Sendo assim, há um amplo campo de estudo a ser revelado sobre esse assunto.

Os resultados mostraram que, na maioria dos estudos encontrados sobre o presenteísmo, há uma amostra voltada para o contexto de trabalho de profissionais da saúde (como enfermeiros e médicos) ou de docentes, havendo, portanto, raros achados sobre o tema com outras amostras (SHIMABUKU; MENDONÇA; FIDELIS, 2017).

Lohaus e Habermann (2019) enfatizaram em seus estudos que, embora a atenção dada ao presenteísmo nos meios acadêmicos e profissionais tenha aumentado, necessita-se da geração de mais conhecimento científico sobre o fenômeno.

Suzuki et al. (2015) propuseram avaliar a influência do presenteísmo na depressão e ausência por doença psicológica. Para isso, eles realizaram um estudo prospectivo de 1.831 empregados japoneses de todas as áreas do Japão. O presenteísmo e a depressão foram medidos pelo Questionário de Desempenho em Saúde e Trabalho da Organização Mundial da Saúde (OMS-HPQ) e pela escala K6, respectivamente.

Os resultados mostraram que mais escores de presenteísmo por doença estão relacionados a maiores taxas de depressão e absenteísmo devido à doença psicológica nesta coorte de larga escala de empregados japoneses. A medição do presenteísmo pode ser usada para avaliar o risco de depressão e absenteísmo. Além disso, os resultados sugerem que a intervenção para melhorar o presenteísmo seria eficaz na prevenção da depressão e absenteísmo devido a doença psicológica (SUZUKI et al., 2015).

Karanika-Murray et al. (2015) propuseram um modelo de mediação do presenteísmo por doença como determinante da satisfação no trabalho por meio de estados emocionais afetivos (especificamente engajamento e dependência ao trabalho). Esse modelo adiciona à literatura atual, focando em satisfação no trabalho como um resultado do presenteísmo, e os processos psicológicos associados a isso. Os resultados da análise de mediação com variáveis observáveis suportaram o modelo.

O presenteísmo por doença foi negativamente associado à satisfação no trabalho. Esse relacionamento foi totalmente mediado tanto pelo envolvimento com o trabalho quanto pelo

vício em trabalhar, chegando a um total de 48,07% da variação no trabalho satisfação. Apesar da pequena amostra, os dados fornecem suporte preliminar para o modelo (KARANIKA-MURRAY et al., 2015).

Os resultados mostraram que não há pesquisas disponíveis sobre as consequências atitudinais do presenteísmo por doença, e essas descobertas são promissoras para o avanço da teorização nessa área. Ao oferecer um modelo que explica os resultados afetivos e motivacionais do presenteísmo, o estudo oferece uma linha de base que pode reforçar ainda mais a compreensão de suas consequências, além de levar a novas e testáveis hipóteses, e apontar para novas direções de pesquisa do presenteísmo por doença (KARANIKA-MURRAY et al., 2015).

Macgregor, Cunningham e Caverley (2008) propuseram investigar a relação de eventos estressantes da vida e eventos relacionados à saúde com absenteísmo e presenteísmo por doença. Os achados indicaram que os eventos estressantes da vida estavam significativamente associados com o absenteísmo e com o presenteísmo no mesmo grau, aumentando ainda mais a evidência de que os fatores que predizem o absenteísmo também predizem o presenteísmo, ou seja, em algumas circunstâncias, os empregados podem substituir o absenteísmo por doença pelo presenteísmo por doença. O estudo enfatiza que o absenteísmo e o presenteísmo são medidas importantes da saúde dos empregados e da produtividade organizacional (MACGREGOR; CUNNINGHAM; CAVERLEY, 2008).

2.3 ANTROPOMETRIA

O estudo do tamanho e da forma do corpo humano tem sido tema de pesquisa há muito tempo (BRAGANÇA et al., 2016). A palavra antropometria é de origem grega (Antropo= homem; metria= medida) e tem por definição o estudo das medidas físicas do corpo humano, sendo, portanto, de vital importância para a ergonomia. A antropometria é fundamental diante do contexto ergonômico, pois é por meio dela que ocorre a adaptação entre os equipamentos, as máquinas e ferramentas e as características físicas do homem (NORTON; OLDS, 2005).

Para Paschoarelli e Santos (2009), a antropometria é a ciência da medição e a arte da aplicação que estabelece a geometria física, as propriedades de massa e as capacidades de força do corpo humano.

De modo geral, a antropometria se refere ao tamanho e às proporções do corpo humano. Toda população é composta de indivíduos de diferentes tipos físicos que apresentam distinções nas proporções de cada segmento do corpo. É comum a antropometria tratar de medidas físicas corporais para verificar seu grau de adequação aos postos de trabalho. As amostragens da

população e as condições em que as medições são realizadas (com roupa ou sem roupa, com ou sem calçados, ereto ou na postura relaxada) influem consideravelmente nos resultados (KROEMER; GRANDJEAN, 2005; SUTALAKSANA; WIDYANTI, 2016; IIDA; BUARQUE, 2016).

A empresa moderna precisa de medidas antropométricas cada vez mais detalhadas e confiáveis. Até a década de 1940, as medidas antropométricas determinavam apenas algumas grandezas médias da população, como peso e estatura. Mais tarde, passaram a determinar também, as dispersões dessas medidas (variância e desvio-padrão) e os alcances dos movimentos.

Atualmente, o principal interesse se concentra no estudo das diferenças entre grupos e a influência de certas variáveis, como etnias, idade, alimentação e saúde. Com o crescente volume do comércio internacional, pensa-se, hoje, em estabelecer os padrões mundiais de medidas antropométricas, possibilitando a produção de produtos universais, adaptáveis aos usuários de diversos países e regiões, mediante a aplicação de métodos de síntese antropométrica, reduz-se o consumo de matéria-prima e aumenta-se vida útil do produto (NADADUR e PARKINSON, 2013; IIDA e BUARQUE, 2016).

Existem três tipos de medidas antropométricas que condicionam as medições e, posteriormente, a aplicação dos resultados obtidos: estática ou estrutural, funcional e dinâmica. Os tipos de medida devem ser selecionados de acordo com o objetivo a ser alcançado (IIDA; BUARQUE, 2016):

a) Antropometria estática ou estrutural

Nessa classe, as medições são realizadas nos segmentos corporais, entre pontos anatômicos claramente identificados, com o corpo parado. Os dados de antropometria estática são recomendados para dimensionar produtos e locais de trabalho, onde ocorrem apenas pequenos movimentos corporais, como no caso do mobiliário em geral.

b) Antropometria funcional

A antropometria funcional é aplicada principalmente quando há uma conjugação de diversos movimentos corporais para a execução de certas tarefas específicas, como acionar uma manivela para fechar o vidro do carro, por exemplo. Na prática, observa-se que cada parte do corpo não se move isoladamente, mas há uma conjugação entre diversos movimentos executados simultaneamente para se realizar uma função. Esses movimentos interagem entre si, modificando os alcances, em relação aos valores da antropometria dinâmica.

c) Antropometria dinâmica

Mede os alcances dos movimentos corporais. Na antropometria dinâmica, as medidas são feitas entre pontos anatômicos, tomados com o sujeito realizando algum movimento. Elas complementam os dados da antropometria estática e contribuem para realizar projetos mais precisos.

Os movimentos de cada segmento corporal são medidos separadamente, mantendo-se o resto do corpo estático. Portanto, não se consideram as interações entre os vários movimentos corporais. Exemplo: alcance máximo das mãos com a pessoa sentada – linha de montagem, caixa de supermercado. Deve-se aplicar a antropometria dinâmica nos casos de trabalhos que exigem muitos movimentos corporais ou quando se deve manipular partes que se movimentam em máquinas ou postos de trabalho.

2.3.1 Pesquisas relacionadas à antropometria

Fonseca e Santos (2013) propuseram avaliar a utilidade da antropometria na identificação e graduação da gravidade da desnutrição, como parte da avaliação inicial de doentes submetidos à gastrostomia endoscópica. Selecionaram-se adultos doentes com registo prévio à gastrostomia: Índice de massa corporal (IMC), perímetro braquial, prega cutânea tricipital e circunferência muscular do braço.

A antropometria na avaliação inicial de doentes gastrostomizados permitiu identificar casos de desnutrição, classificar sua gravidade e estimar o seu impacto. Perímetro braquial, prega cutânea tricipital e circunferência muscular do braço identificaram desnutrição em cerca de 80% dos doentes, mais na massa adiposa do que na muscular. Acessível e não dispendiosa, a antropometria permitiu adequar o suporte nutricional (FONSECA; SANTOS, 2013).

Os resultados mostraram que a obtenção dos dados antropométricos individuais foi possível em todos os doentes e exigiu apenas recursos materiais de baixo custo, conhecimentos anatômicos coerentes e rigor técnico dos profissionais de saúde, mas, em contrapartida, os dados antropométricos permitiram o planejamento do suporte nutricional individual, quer quantitativamente, quer do ponto de vista qualitativo (FONSECA e SANTOS, 2013).

Mesquita e Mesquita (2013) propuseram determinar o perfil antropométrico, socioeconômico e de saúde de 103 empregados de duas empresas do ramo industrial de confecção, assistidos pelo Programa de Alimentação do Trabalhador (PAT) em Fortaleza, Ceará, Brasil, e verificar a adequação calórica na alimentação recebida em seu ambiente de trabalho. Os empregados avaliados tinham um índice de massa corporal (IMC) de sobrepeso, estavam em tratamento de doenças crônicas não degenerativas e eram, em sua maioria,

mulheres com renda de até um salário mínimo, cuja alimentação recebida era inadequada em calorias, carboidratos e proteínas.

Schlosser, Debiassi e Parcianello (2002) propuseram determinar o padrão antropométrico dos operadores de tratores agrícolas da Depressão Central do Rio Grande do Sul. As variáveis medidas foram: altura do corpo; altura ao nível dos olhos; altura ao nível dos olhos sentado; altura do cotovelo; alcance do braço; alcance da mão; distância entre o pé e a patela; e apoio do assento. Os dados obtidos demonstraram que há diferenças entre o biótipo do operador utilizado pela indústria de tratores agrícolas e o do operador da região. Os padrões determinados pelo estudo levaram em consideração as diferenças existentes entre o perfil antropométrico dos operadores de tratores e os parâmetros utilizados pela indústria. Com isso, confirmou-se a hipótese de que os tratores agrícolas que se encontram atualmente em comercialização no Brasil podem não oferecer o conforto necessário ao operador dessa região.

Os resultados também mostraram que o padrão antropométrico dos operadores de tratores agrícolas da região da Depressão Central do Rio Grande do Sul é comparativamente diferente ao padrão dos países desenvolvidos; portanto, há necessidade de algumas modificações no projeto dos tratores, em relação àqueles adequados aos operadores de países desenvolvidos (SCHLOSSER; DEBIASSI; PARCIANELLO, 2002).

Ochoa et al. (2009) propuseram identificar, por meio da antropometria adequada para polo aquático, e usando mineração de dados sociais, as diferenças nos parâmetros antropométricos, isto é, a gordura corporal, o índice de massa corporal (IMC) e a densidade corporal induzida pela otimização morfológica específica do esporte (adaptação) entre duas amostras (Universidade Juarez City e Universidade Tecnológica de Montenegro) de 160 jogadores masculinos de polo aquático.

Antropometristas treinados e qualificados realizaram a medição em condições experimentais padronizadas e de acordo com os procedimentos descritos pelo Programa Biológico Internacional. Eles mediram 23 variáveis antropométricas descritas pelos comprimentos ósseos esqueléticos, pelas circunferências (peito, braço, antebraço, coxa e perímetro da panturrilha), pela espessura das dobras cutâneas (tríceps, subescapular, axilar, panturrilha e pele abdominal), e pela espessura da dobra e massa. Além disso, as estimativas do índice de massa corporal (IMC), da densidade corporal e da porcentagem de gordura corporal foram calculadas a partir das medidas primárias (OCHOA et al., 2009).

Os resultados forneceram uma boa visão sobre as alterações antropométricas nos jogadores de polo aquático. Esse estudo pode ser um desafio para uma investigação mais

aprofundada dos fenômenos de tendência secular da população e de otimização morfológica do esporte, com atenção especial dada à sua possível interação (OCHOA et al., 2009).

2.4 ERGONOMIA

A ergonomia, por definição, é o estudo da adaptação do trabalho ao ser humano. O trabalho abrange não apenas os trabalhos executados com máquinas e equipamentos, utilizados para transformar os materiais, mas também todas as situações em que ocorre o relacionamento entre o ser humano e uma atividade produtiva de bens ou serviços, envolvendo o ambiente físico e os aspectos organizacionais necessários para que o trabalho possa atingir os resultados esperados, como (IIDA; BUARQUE, 2016):

- a) Planejamento e projeto: ocorrem antes do trabalho a ser realizado;
- b) Monitoramento, avaliação e correção: ocorrem durante a execução desse trabalho e;
- c) Análises posteriores das consequências do trabalho.

São estudados na ergonomia vários aspectos de adaptação do trabalho ao homem e do comportamento humano no trabalho, tais como o ser humano, a máquina e o ambiente de trabalho. Na ergonomia, aborda-se também o conhecimento de diferentes áreas científicas, como a antropometria, a biomecânica, a toxicologia, e as engenharias, bem como a aplicação dos conhecimentos de anatomia, fisiologia e psicologia, para a solução dos problemas surgidos desse relacionamento (DUL; WEERDMEESTER, 2004; IIDA; BUARQUE, 2016).

De acordo com Wisner (1992), a ergonomia, na área do trabalho, apresenta um grande desafio, que é adaptar, seja na concepção ou na correção, os postos de trabalhos de acordo com as características físicas das pessoas; o resultado das medidas antropométricas deve ser capaz de proporcionar maior conforto, saúde e segurança, possibilitando uma postura adequada.

A ergonomia tem por propósito, conforme determinados procedimentos referentes à sua utilização, melhorar a qualidade de vida das pessoas nos mais diferentes ambientes, essencialmente, em locais de trabalho, proporcionando conforto e segurança (IIDA; BUARQUE, 2016).

As variadas definições existentes na literatura conceituam a ergonomia como uma ciência, associando-a a diversos enfoques. A ergonomia é o conjunto dos conhecimentos científicos relativos ao homem, que é necessário para a criação de ferramentas, máquinas e dispositivos que possam ser utilizados com maior conforto, segurança e eficácia. Essa ciência utiliza dados antropométricos, fisiológicos e psicológicos para a concepção e melhoria do dispositivo de trabalho (WISNER, 1992; PIZO; MENEGON, 2010).

No entanto, a definição oficial de ergonomia foi adotada pelo *International Ergonomics Association* (IEA), em agosto de 2000, como uma disciplina científica relacionada ao entendimento das interações entre o homem e outros elementos de um sistema e à aplicação de teorias, princípios, dados e métodos, a fim de aperfeiçoar o desempenho global do sistema (ABERGO, 2020).

2.4.1 Medições em ergonomia

Segundo Iida e Albuquerque (2016), as pesquisas em ergonomia realizam vários tipos de medições das variáveis envolvidas no sistema humano-máquina-ambiente. Quando se trata de medir certa variável material, como a altura de uma cadeira ou mesa, o problema é relativamente simples, contudo, para obter dados relativos à parte humana do sistema, a tarefa torna-se mais complexa e requer uma série de cuidados para evitar ou reduzir os erros.

Para tanto, as medições do ser humano devem ser realizadas em condições padronizadas, que constem de instruções escritas. Por exemplo, a medida da estatura de uma pessoa pode variar em até 5 cm, dependendo do horário em que é realizada a medição e de como a pessoa está, isto é, com ou sem calçado, na posição ereta ou na postura natural, e assim por diante (IIDA; ALBUQUERQUE, 2016).

As medições do ser humano recaem em duas categorias gerais: objetivas e subjetivas.

a) Medições objetivas: são aquelas realizadas com o auxílio de instrumentos de medida e resultam em um determinado valor numérico.

Um exemplo de medidas objetivas são as antropométricas e biomecânicas, como estatura, peso e força. Também se incluem as medidas fisiológicas, como a temperatura corporal, a composição do sangue e da urina e as atividades elétricas da musculatura. Deve-se tomar certo cuidado na hora escolher as variáveis e de definir a quantidade de medições, a faixa de valores, o tipo de instrumento a ser utilizado (isso influi na precisão) e as condições em que serão feitas as medições (IIDA; ALBUQUERQUE, 2016).

b) Medições subjetivas: são aquelas que dependem de algum tipo de julgamento dos sujeitos.

Por exemplo, fadiga e conforto dependem de muitos fatores e dificilmente podem ser determinados por medições instrumentais, ainda que indiretamente. Nesse caso, o sentimento de fadiga ou conforto deve ser manifestado pelo sujeito. Isso pode levar a erros experimentais, mas que podem ser reduzidos por um planejamento e controle adequado do experimento. As medições subjetivas nem sempre podem ser quantificadas em números, mas podem ser qualificadas ou classificadas. Elas são baseadas geralmente em entrevistas e questionários. (IIDA; ALBUQUERQUE, 2016).

2.4.2 Especialização da ergonomia

Segundo a Associação Brasileira de Ergonomia (ABERGO, 2020), os domínios de especialização da ergonomia estão divididos em:

a) Ergonomia física

Está associada às características da anatomia humana, antropometria, fisiologia e biomecânica em sua relação com a atividade física. Os tópicos pertinentes envolvem o estudo da postura no trabalho, do manuseio de materiais, dos movimentos repetitivos, dos distúrbios músculo-esqueléticos referentes ao trabalho, do projeto de posto de trabalho, da segurança e da saúde.

b) Ergonomia cognitiva

Está associada aos processos mentais, como percepção, memória, raciocínio e controle motor, e aos seus efeitos nas interações entre seres humanos e outros elementos de um sistema. Os tópicos pertinentes incluem o estudo da carga mental de trabalho, da tomada de decisão, do desempenho especializado, da interação entre homem e computador, do estresse e do treinamento, adaptando-os na concepção da relação entre homem e sistema.

c) Ergonomia organizacional

Está relacionada à otimização dos sistemas sociotécnicos, incluindo sua estrutura organizacional, suas políticas e seus processos. Associa-se às comunicações, ao gerenciamento de recursos de tripulações (domínio aeronáutico), ao projeto de trabalho, à organização temporal do trabalho, ao trabalho em grupo, ao projeto participativo, aos novos paradigmas do trabalho, ao trabalho cooperativo, à cultura organizacional, às organizações em rede, ao teletrabalho e à gestão da qualidade (ABERGO, 2020).

Foi a partir do 1º Seminário Brasileiro de Ergonomia, em 1975, que a contribuição da ergonomia de Alain Wisner, pesquisador do Laboratório de Ergonomia e Neurociências do Trabalho (CNAM - *Conservatoire National des Arts et Métiers*), na França, foi decisiva para o processo de institucionalização da ergonomia e para o desenvolvimento da Análise Ergonômica do Trabalho (AET), no Brasil. Desde então, vários pesquisadores e agentes públicos brasileiros se formaram no Laboratório de Ergonomia (CNAM), dirigido por Alain Wisner (JACKSON FILHO; LIMA, 2015).

De qualquer forma, quanto mais a competência ergonômica se aproxima das pressões da transformação, mais ela deve posicionar-se para a aprendizagem das atividades, dos valores e dos saberes (PIZO; MENEGON, 2010).

2.4.3 Pesquisas relacionadas à ergonomia

É importante ressaltar que, segundo Jackson Filho e Lima (2015), a ergonomia tem sido muito divulgada no Brasil, principalmente por meio de artigos publicados em revistas nacionais sobre os mais variados assuntos referentes à ergonomia:

Batiz, Vergara e Licea (2012) apresentaram resultados da aplicação dos métodos, dois relacionados ao carregamento de carga e outro de análise postural, a fim de verificar a existência de riscos de problemas dorsolombares (porção lombar da coluna vertebral) em auxiliares de enfermagem na manipulação e no carregamento de pacientes. Também aplicaram o método NIOSH, um guia técnico para avaliação e prevenção dos riscos relativos à manipulação manual de cargas, e o método REBA (*Rapid Entire Body Assessment*), que é um método de análise postural adotado pelos auxiliares de enfermagem para conhecer os possíveis riscos de danos à saúde.

Os resultados mostraram a existência de sintomas de fadiga e dores no corpo, principalmente nas regiões lombar, dorsal e cervical dos auxiliares de enfermagem, e constataram que o peso manipulado está acima do valor recomendado pelos dois métodos aplicados, caracterizando alto risco para a região dorsolombar. Considerando que a atividade de manipulação e carregamento de pacientes constitui por si só um fator de risco dorsolombar para os auxiliares de enfermagem, os autores esperam que pesquisa possa servir de subsídio para o trabalho preventivo, junto aos auxiliares de enfermagem, em prol de melhores condições de trabalho e, conseqüentemente, melhor qualidade de vida e saúde para esses profissionais (BATIZ; VERGARA; LICEA (2012).

Rothstein et al. (2013) avaliaram o impacto de uma metodologia interativa de ergonomia de conscientização no conhecimento de empregados de uma indústria têxtil. A amostra foi selecionada a partir do parecer do comitê de ergonomia, e participaram do estudo 328 empregados. Foram aplicadas intervenções interativas nos seguintes temas: corpo humano, visão, coluna vertebral, membros superiores e membros inferiores. O impacto da metodologia foi mensurado com a aplicação de um questionário criado e validado para o estudo.

Os resultados apontaram a efetividade da metodologia quanto ao nível de conhecimento. Com esse estudo, concluiu-se que informar o empregado da indústria sobre segurança e saúde, ergonomia e consciência corporal, com suporte de recursos didáticos interativos e lúdicos, que contemplam a realidade do indivíduo, pode ser uma alternativa promissora na prevenção de doenças ocupacionais, visto que é necessário sensibilizá-los para a questão do autocuidado, dentro e fora do ambiente de trabalho (ROTHSTEIN et al., 2013).

Pinto, Tereso e Abrahão (2015) sugeriram identificar e compreender as práticas ergonômicas adotadas nas indústrias da Região Metropolitana de Campinas de acordo com a sua natureza, a gestão e os atores sociais envolvidos. Foram aplicados questionários para identificar quais as indústrias que possuíam análises e práticas ergonômicas.

Os resultados mostraram que as práticas ergonômicas favorecem a compreensão da atividade, dando significância ao trabalho, estabelecem uma interlocução entre os atores envolvidos nos diferentes níveis hierárquicos, contribuem para as transformações e melhorias, no sentido de preservar a saúde e segurança dos empregados, e cooperam para um melhor desempenho da organização (PINTO; TERESO; ABRAHÃO (2015).

Ferreira (2015) apresentou um relato sobre uma experiência de uso heterodoxo da Análise Ergonômica do Trabalho (AET), situando seu papel instrumental no nível analítico microergonômico da abordagem intitulada Ergonomia da Atividade Aplicada à Qualidade de Vida no Trabalho (EAA_QVT).

Os resultados produzidos pela aplicação da EAA_QVT revelou a possibilidade de gerar recomendações para as atividades de vigilância, assistência e promoção da saúde e segurança no trabalho. O uso da EAA_QVT revelou ainda certos limites e, paulatinamente, vem acumulando experiências de sua aplicação, ainda restritas ao setor público brasileiro. A perspectiva de aplicação da EAA_QVT, situando o lugar, a importância e a contribuição da AET, poderá gerar, nas organizações, fundamentos empíricos e teóricos para a formulação de políticas e de programas de QVT, de modo a alterar positivamente o cenário atual do mundo do trabalho (FERREIRA, 2015).

Santos et al. (2016) apresentaram uma análise de um sistema técnico, formado pela integração de um sistema de captura de movimentos, que inclui manequins digitais com um software de modelagem e simulação humana, ou seja, criaram procedimentos que relacionam as interações possíveis entre o corpo de captura e o ambiente de trabalho com as características da atividade e os objetivos da simulação, visando estabelecer formas de utilização no processo de projetos contextualizados pela ergonomia. Os autores propuseram um sistema integrado composto por duas diferentes tecnologias computacionais, o sistema de MoCap Moven, que utiliza sensores inerciais, e o software CAD Delmia, que possui um módulo de modelagem e simulação de movimento humano digital.

Os resultados do estudo mostraram que a utilização de um sistema integrado permite capturar movimentos reais, realizados pelos empregados ou usuários, e posicioná-los no ambiente virtual, referenciando o local real de trabalho. Além disso, o sistema contribui para redução do tempo de criação desse tipo de análise e para a melhoria dos erros de construção

postural e de posicionamento associado ao método, que possibilita uma melhor incorporação da perspectiva da atividade conforme o conceito da ergonomia situada (SANTOS et al., 2016).

Pimentel et al. (2017) propuseram investigar os efeitos da ginástica laboral sobre determinantes de desconforto e dores musculoesqueléticas enquanto indicativo sintomatológico relacionado ao desenvolvimento de lesões por esforços repetitivos e distúrbios osteomusculares relativos ao trabalho, e isso por meio de uma revisão sistematizada da literatura científica entre os anos de 2006 a 2015.

Os estudos mostraram que, mesmo que a prática da ginástica laboral apresente resultados favoráveis à redução de sintomatologias que se associam a indicativos de problemas relacionados a distúrbios osteomusculares, sua intervenção ainda é limitada devido à complexidade do trabalho. Neste sentido, torna-se viável pensar na ginástica laboral, enquanto uma estratégia, junto a um programa de gestão de saúde e qualidade de vida no trabalho. Além disso, é possível estabelecer ações de forma a contribuir para os objetivos da prevenção das DORT. Os autores alertam para um número reduzido de estudos validados para esta pesquisa, mediante um tema que permuta as dimensões afirmativas da contribuição da ginástica laboral (PIMENTEL et al., 2017).

Souza e Camarotto (2017) relacionaram as tecnologias na coleta seletiva com a perspectiva da antropotecnologia, termo que refere à macroergonomia utilizada por Hendrick. Toda máquina é cultural, e a passagem de uma tecnologia à outra é difícil. A partir dessas ideias centrais, é que se desenvolve a antropotecnologia, que visa reorientar as ciências do homem coletivo a fim de agir sobre as exigências materiais e as características etnológicas. Seu campo é a transferência de tecnologia e, seu objetivo, modificar os sistemas técnicos e organizacionais (WISNER, 1992).

Os resultados nos estudos de Souza e Camarotto (2017) mostram que a ergonomia, com base na antropotecnologia, é uma importante ferramenta para a compreensão dos conjuntos de técnicas utilizadas pelos empregados e de como eles se organizam em função dessas tecnologias.

Pizo e Menegon (2010) propuseram contextualizar, na Análise Ergonômica do Trabalho (AET), as preocupações de pesquisadores da área de ergonomia sobre o reconhecimento científico dos métodos de pesquisa sobre o trabalho. Ambos também propuseram uma análise teórica da simbiose entre esse método de pesquisa e a AET e de como isso contribui para que os conhecimentos gerados possam ter seu reconhecimento científico facilitado junto aos pares das áreas de conhecimento em que a ergonomia permeia.

Os resultados dos estudos mostraram que para uma pesquisa ser reconhecida é preciso, antes, validá-la perante a comunidade em que ela é apresentada, e que sua aceitação depende em parte do método utilizado e de sua condução. Por isso é importante que a condução da pesquisa e a representação do conhecimento sejam aceitas pela comunidade na qual se quer validá-la. Os resultados mostraram também que a aceitação não é estática, nem unânime e se altera com o tempo, conforme a evolução da área do conhecimento.

Assim, os trabalhos desenvolvidos devem ser sempre submetidos à comunidade para descobrir como está evoluindo o seu pensar sobre o tema, o método, e também para modificar esse caminho (PIZO; MENEGON, 2010).

2.5 MINERAÇÃO DE DADOS

O termo Mineração de Dados, ou Data Mining, surgiu de uma analogia entre a procura de informação, importante numa base de dados, e o ato de minerar a montanha em busca de um veio de ouro. A mineração de dados é o elemento responsável pela extração eficiente do conhecimento implícito e útil contido em um banco de dados (BIGUS, 1996).

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, utilizando uma série de algoritmos inteligentes para encontrar relações fundamentais entre os dados. Essas técnicas e algoritmos se caracterizam por serem automáticos. A participação humana se restringe aos processos de ajustes necessários aos mecanismos de busca (PIATETSKY-SHAPIRO; MATHEUS; CHAN, 1993; FONSECA; NAMEN, 2016).

Uma das vantagens de se utilizar as técnicas de mineração de dados é que elas permitem avaliar como as perguntas se relacionam com as respostas, com os padrões e com as relações encontradas. Para tal, utilizam-se técnicas baseadas em IA, como as Redes Neurais Artificiais (RNAs), as árvores de decisões, a teoria dos conjuntos *Fuzzy*, os Algoritmos Genéticos ou, ainda, combinações entre essas técnicas, gerando os chamados Sistemas Híbridos ou Arquiteturas Híbridas (HAN; KAMBER; PEI, 2012).

Segundo Silva, Peres e Boscaroli (2017), a partir dos padrões descobertos, têm-se condições de gerar conhecimento útil para auxiliar numa tomada de decisão. Trata-se da aplicação de técnicas implementadas por meio de algoritmos computacionais capazes de receber, como entrada, um conjunto de fatos ocorridos no mundo real, e de devolver, como

saída, um padrão de comportamento que pode ser expresso, por exemplo, como uma regra de associação, uma função do mapeamento ou da modelagem.

Na Figura 1, mostram-se as etapas do processo simplificado de mineração de dados.

Figura 1 – Etapas do processo simplificado de mineração de dados



Fonte: Adaptado de Bigus (1996) e Han, Kamber e Pei (2012).

Conforme Bigus (1996), o processo de mineração de dados consiste em três etapas principais. A primeira etapa do processamento é a preparação, em que os dados são selecionados, limpos e pré-processados quando possível sob a orientação e o conhecimento de um especialista no domínio dos dados.

Na segunda etapa, um algoritmo de mineração de dados é usado para processar os dados preparados, compactando-os e transformando-os para facilitar a identificação de qualquer conhecimento valioso não aparente.

A terceira etapa se refere à descoberta do conhecimento ou análise de dados, na qual a saída da mineração de dados é avaliada para verificar se o conhecimento adicional foi descoberto e para determinar a importância relativa dos fatos gerados pelos algoritmos de mineração. É nele que as decisões são tomadas, usando as informações coletadas pelo processo de mineração de dados, e que o conhecimento é aplicado para determinado objetivo.

O resultado obtido pela aplicação da mineração de dados deve ser compacto, legível e apresentado de forma interpretável, representando fielmente os dados que lhe deram origem (BACK, 2002).

A etapa final do processo de mineração consiste no pós-processamento, que engloba a interpretação dos padrões descobertos e a possibilidade de retorno a qualquer um dos passos anteriores. Assim, a informação extraída é analisada ou interpretada em relação ao objetivo proposto, sendo identificados e apresentados os melhores resultados das informações (FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996).

2.5.1 Tarefas de mineração de dados

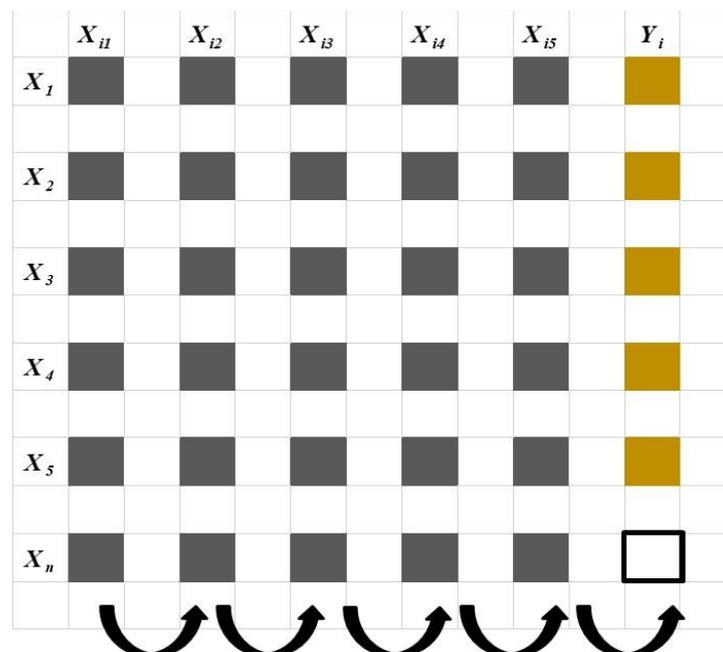
A mineração de dados é comumente considerada pela sua capacidade em realizar determinadas tarefas, e os algoritmos são utilizados de acordo com os objetivos do estudo, a fim de se obter uma resposta para o problema (LAROSE; LAROSE, 2014). As tarefas mais comuns são (CARDOSO; MACHADO, 2008):

a) Classificação

Na tarefa de classificação, geram-se modelos que descrevem e distinguem classes ou conceitos, baseados em dados conhecidos, com o propósito de utilizar esse modelo para prever a classe de objetos que ainda não foram classificados (CARDOSO; MACHADO, 2008).

A tarefa de classificação consiste na análise de um conjunto de dados nos quais estão presentes os dados descritos por atributos e seus rótulos associados. O objetivo, nessa tarefa, é descobrir um modelo capaz de mapear corretamente cada um dos dados X_1, X_2, X_3, X_4, X_5 aos seus rótulos Y . Na Figura 2, ilustra-se a tarefa de classificação na mineração de dados.

Figura 2 – Tarefa de classificação



Fonte: Elaborado pelo autor (2020).

De acordo com Farid et al. (2014), essa tarefa consiste em uma função de aprendizado que mapeia dados de entrada, ou conjunto de dados de entrada, em um número finito de classes. O processo se inicia com a divisão do conjunto de dados em dois totalmente distintos: o conjunto de treinamento e o conjunto de testes, a fim de encontrar uma função que associe um caso a uma classe dentro de diversas classes discretas de classificação.

As técnicas de classificação criam automaticamente um modelo, que serve de exemplo, a partir de um conjunto inicial de registros, sendo chamado de conjunto de treinamento. Os registros do conjunto de treinamento devem pertencer a um pequeno grupo de classes pré-definidas pelo analista (HAN; KAMBER; PEI, 2012).

Assim, refinando a definição da tarefa de classificação, diz-se que é o processo pelo qual se determina um mapeamento capaz de indicar a qual classe pertence qualquer exemplar de um domínio sob análise, com base em um conjunto de dados já classificado (SILVA; PERES; BOSCARIOLI, 2016).

Na mineração de dados, a tarefa de previsão consiste em prever uma variável categórica em um conjunto de registros predefinidos, sendo denominados como classes. Tal tarefa pode ser aplicada em novos registros, de maneira a prever sua classe (KESAVARAJ; SUKUMARAN, 2013).

b) Agrupamento

O agrupamento, também chamado de clusterização, é uma tarefa em que se busca determinar grupos ou segmentos para descrever os dados. Isso consiste na identificação de grupos semelhantes de objetos em que cada grupo é uma classe. Dentro da mesma classe, os objetos são semelhantes, e, entre as classes, são divergentes (FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996).

O agrupamento transforma registros com grande número de atributos em conjuntos relativamente menores. Essa transformação é realizada, automaticamente, por meio da identificação das características que distinguem o conjunto de dados e pelo seu posterior particionamento (BRAGA; CARVALHO; LUDERMIR, 2011).

Diferente do processo de classificação, no agrupamento, não há a presença de um rótulo associado. As análises realizadas pelos algoritmos, que implementam estratégias para agrupamento, buscam por similaridades ou diferenças entre exemplares, quantificadas por meio de medidas de distância; quanto maior for a distância entre dois exemplares, maior será a similaridade (SILVA; PERES; BOSCARIOLI, 2016).

A tarefa de agrupamento não necessita que os registros sejam previamente categorizados, ou seja, o aprendizado não é supervisionado. No aprendizado não-supervisionado, não há um professor para supervisionar o processo de aprendizagem (HAYKIN, 2001).

Na Figura 3, ilustra-se a tarefa de agrupamento na mineração de dados.

Figura 3 – Tarefa de agrupamento

	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	Y_i
X_1	■	■	■	■	■	■
X_2	■	■	■	■	■	■
X_3	■	■	■	■	■	■
X_4	■	■	■	■	■	■
X_5	■	■	■	■	■	■
X_n	■	■	■	■	■	□

Fonte: Elaborado pelo autor (2020).

O objetivo dessa tarefa é descobrir as relações, de similaridades ou diferenças, dos dados $X_1, X_2, X_3, X_4, X_5, X_n$, e fornecer, como resposta, a indicação de quais dados são similares entre si, oferecendo um modelo de agrupamento ou perfis para o grupo de dados (SILVA; PERES; BOSCARIOLI, 2016).

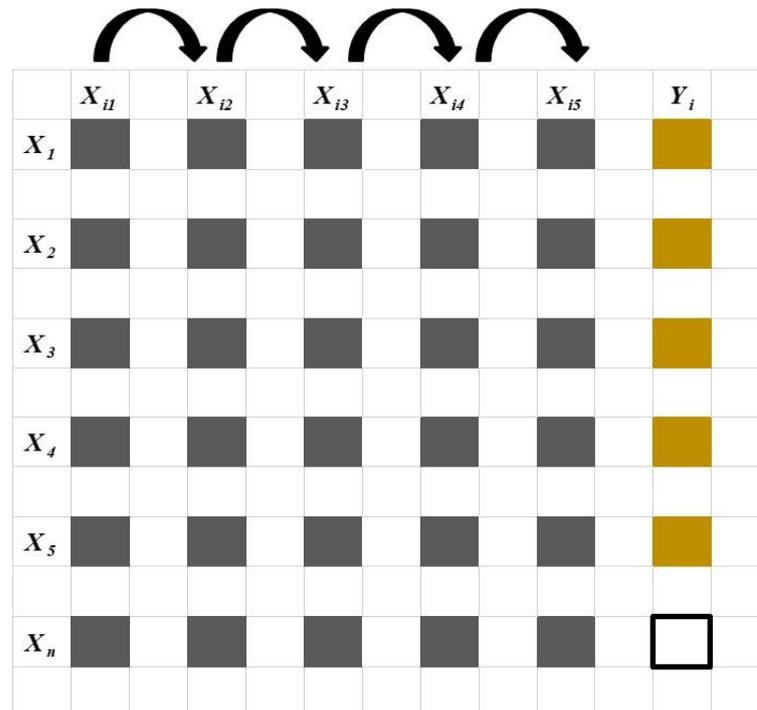
c) Associação

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 (SADATH, 2013).

A tarefa de associação tem por definição a busca por ocorrências frequentes e simultâneas entre elementos de um contexto. Os algoritmos que resolvem essa tarefa analisam conjuntos de dados que apresentem eventos ou transações, procurando por itens frequentemente envolvidos nos mesmos eventos ou que apresentam algum tipo de correlação em seus comportamentos em tais eventos.

Na Figura 4, ilustra-se a tarefa de associação na mineração de dados.

Figura 4 – Tarefa de associação



Fonte: Elaborado pelo autor (2020).

Nesse tipo de tarefa, deve-se considerar, como importantes, não apenas as associações triviais, mas também aquelas não tão óbvias, uma vez que se transformam em uma relevante fonte de informação na tomada de decisão (GOLDSCHMIDT; PASSOS; BEZERRA, 2015; WITTEN et al., 2016).

2.5.2 Pesquisas relacionadas à mineração de dados

A mineração de dados é aplicada nas mais diversas áreas do conhecimento, seja para a resolução de problemas, de gerenciamento de recursos humanos e de talentos, seja como ferramenta para gerenciamento do conhecimento em recursos humanos (STROHMEIER; PIAZZA (2013). Jantan, Hamdan e Othman (2009) aplicaram a mineração em informações oriundas de prontuários de pacientes; Sadath (2013) aplicou mineração para identificar talentos organizacionais; Carvalho (2018) aplicou na educação em engenharia; Blum, Merino e Merino (2016), em problemas de dados faltantes em pesquisas clínicas; Leandro et al. (2018), em avaliação de óbitos após cirurgia de amputação; e Rabbi et al. (2018) aplicaram em base de reclamações sobre produtos e serviços do programa de proteção e defesa do consumidor.

Cardoso e Machado (2008) propuseram desenvolver, aplicar e analisar uma ferramenta de mineração de dados, para extrair conhecimento, referente à produção científica dos envolvidos com a pesquisa na Universidade Federal de Lavras (UFLA), por meio dos dados

extraídos da plataforma Lattes. Foi implementado um programa para transformar os dados semiestruturados selecionados dessa plataforma em um banco de dados estruturado criado no Oracle, um sistema gerenciador de banco de dados. A partir daí, foi desenvolvida uma ferramenta automática de descoberta de conhecimento, utilizando a técnica de mineração de dados, cujos resultados gerados foram analisados.

A contribuição do estudo foi gerar um banco de dados estruturado, que faz parte de um processo maior de desenvolvimento de indicadores de ciência e tecnologia, para auxiliar na elaboração de novas políticas de gestão científica e tecnológica e no aperfeiçoamento do sistema de ensino superior brasileiro. Cardoso e Machado (2008) enfatizaram que, apesar das diversas limitações encontradas na gestão dos sistemas de informação, foi possível utilizar, em seu estudo, a técnica de mineração de dados para extrair conhecimento e contribuir para a melhoria do preenchimento dos dados na plataforma Lattes.

Os resultados mostraram que, com essa ferramenta de mineração de dados, é possível obter uma visão mais abrangente dos dados institucionais, pelo fato de ter sido disponibilizada uma grande quantidade de informações sobre a pesquisa científica da UFLA (CARDOSO; MACHADO, 2008).

Carvalho et al. (2012) propuseram apresentar e discutir o potencial de utilização da mineração de dados sobre um conjunto de dados de acompanhamento fisioterapêutico de pacientes, e sua utilidade na tomada de decisões terapêuticas ou profiláticas. Os autores selecionaram um subconjunto de dados, referentes a prontuários disponíveis em uma clínica de fisioterapia, do qual foram extraídos três grandes grupos-alvo de tarefas de mineração de dados: associação, classificação e agrupamento.

Os resultados mostraram que foram extraídos padrões a partir dos dados de tal forma que permitiu ao leitor entender passo a passo o processo, ampliando sua compreensão dos resultados obtidos. Além disso, foram descobertos padrões em diversos formatos, que evidenciaram as possíveis relações entre as variáveis disponíveis. Segundo Carvalho et al. (2012), o estudo permitiu discutir não só os padrões, mas também a importância da qualidade dos dados coletados. Os autores concluíram que as etapas de classificação, descoberta de regras de associação e agrupamento dos dados proporcionaram um melhor entendimento das especificidades de pacientes atendidos pela clínica em questão, ampliando, assim, o conhecimento do profissional na identificação das condutas a serem adotadas (CARVALHO et al., 2012).

Kavakiotis et al. (2017) propuseram conduzir uma revisão sistemática das aplicações de aprendizado de máquina, técnicas e ferramentas de mineração de dados no campo da pesquisa

em Diabetes Mellitus, com relação à predição e ao diagnóstico, às complicações diabéticas, aos antecedentes genéticos, ao ambiente e aos cuidados, e ao gerenciamento de saúde. Foram utilizados conjuntos de dados clínicos. As aplicações dos títulos nos artigos selecionados projetaram a utilidade de se extrair informações valiosas para novas hipóteses, visando um entendimento e uma investigação mais aprofundada em mineração de dados.

Neste estudo, foi feito um levantamento sistemático para identificar e revisar as abordagens de aprendizado de máquina e mineração de dados aplicadas na pesquisa de Diabetes Mellitus. Os resultados do estudo mostraram que o Diabetes Mellitus está emergindo rapidamente como um dos maiores desafios globais de saúde do século XXI.

Kavakiotis et al. (2017) buscam, com os resultados de seus estudos, que o advento da biotecnologia, com a vasta quantidade de dados produzidos, juntamente com a quantidade crescente de grandes registros eletrônicos de saúde, dê origem a uma exploração mais aprofundada em direção ao diagnóstico e tratamento do diabetes por meio de aplicações de aprendizado de máquina e técnicas e de ferramentas de mineração de dados em conjunto de dados enriquecidos, que incluem informações clínicas e biológicas.

Gualhano, Salles e Hora (2018) propuseram utilizar técnicas de mineração de dados para extrair conhecimento da avaliação quadrienal (2013-2016) e identificar os principais critérios na avaliação dos programas de pós-graduação. Para isso, foram coletados dados de avaliações da Plataforma Sucupira e realizado o tratamento dessas informações, utilizando o software WEKA e o método de classificação J48. Buscou-se diagnosticar três desfechos: nota final, evolução da nota e situação do programa. Ainda de acordo com os autores, o estudo foi considerado satisfatório por permitir realizar a classificação dos programas quanto ao desempenho nos itens de avaliação da CAPES e quanto à identificação dos critérios de avaliação com maior impacto nos desfechos analisados.

Os resultados mostraram que tanto o documento de área quanto o regulamento da avaliação quadrienal forneceram os critérios para progressão da nota de avaliação dos programas de PPGs, baseados nos quesitos da ficha de avaliação. Contudo, estes não avançam no detalhamento dos itens de cada quesito. Os resultados também mostraram que a produção intelectual de um PPG é item relevante para manutenção do credenciamento dos programas, bem como para melhoria de sua avaliação. Essa pesquisa atingiu seus objetivos quanto à aplicação de técnicas da mineração de dados, para a classificação dos programas de mestrado acadêmico/profissional e doutorado acadêmico, e quanto ao seu desempenho perante os critérios de avaliação estabelecidos.

2.6 TÉCNICAS DE MINERAÇÃO DE DADOS

Dentre as várias técnicas utilizadas em mineração de dados, destacam-se as Redes Neurais Artificiais (RNAs) com suas diversas arquiteturas: *Random Forest* (RF), ou Florestas Aleatórias, e o Algoritmo de Otimização por Colônia de Formigas (ACO), que vem sendo usado para a tarefa de classificação em mineração de dados (DE AVELAR; ROCHA; CRUZ, 2017).

2.6.1 Redes Neurais Artificiais (RNAs)

As Redes Neurais Artificiais (RNAs) são modelos constituídos por unidades de processamento simples, chamados de neurônios artificiais, que calculam funções matemáticas. Esses modelos são inspirados na estrutura do cérebro e têm como objetivo simular o comportamento humano, como: aprendizagem, associação, generalização e abstração, quando submetidas a treinamento (HAYKIN, 2001).

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, como: medicina (BLAZADONAKIS; MICHALIS, 2008), previsão de risco de crédito (SELAU; RIBEIRO, 2009), data mining (Bigus, 1996), e reconhecimento de padrões (HAYKIN, 2001).

Entre as principais características envolvidas com a aplicação das RNAs, segundo Haykin (2008), estão: adaptação por experiência, capacidade de aprendizado, habilidade de generalização, organização de dados, tolerância a falhas e armazenamento distribuído. Tais características são descritas a seguir:

a) Adaptação por experiência

As adaptações dos parâmetros internos da rede, pesos sinápticos, são ajustadas a partir da apresentação sucessiva de exemplos (padrões, amostras, medidas) relacionados ao comportamento do processo, possibilitando que a rede aprenda por experimentação.

b) Capacidade de aprendizado

Por intermédio da aplicação de um método de treinamento, a rede consegue extrair o relacionamento existente entre as diversas variáveis que compõem a aplicação.

c) Habilidade de generalização

Após o processo de treinamento da rede, é possível generalizar o conhecimento adquirido, estimando soluções até então desconhecidas.

d) Organização de dados

Baseadas em características intrínsecas, envolvendo determinado conjunto de informações a respeito de um processo, a rede é capaz de realizar a sua organização interna, possibilitando o agrupamento de padrões que apresentam particularidades em comum.

e) Tolerância a falhas

Devido ao elevado nível de interconexões entre os neurônios artificiais, a rede neural se torna um sistema tolerante a falhas; quando parte de sua estrutura interna, é sensivelmente corrompida.

f) Armazenamento distribuído:

O conhecimento a respeito do comportamento de determinado processo dentro de uma arquitetura neural é realizado de forma distribuída entre as diversas sinapses de seus neurônios artificiais, permitindo então um incremento da robustez da arquitetura frente a eventuais neurônios que se tornaram inoperantes.

Dentre as várias RNAs disponíveis, e que podem ser aplicadas na solução de diversos problemas, destacam-se:

- A *MultiLayer Perceptron* (MLP), que consiste em um conjunto de 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. Trata-se de uma RNA com aprendizagem supervisionada usada para a tarefa de classificação (SILVA; SPATTI; FLAUZINO, 2016);

- A *Radial Basis Function* (RBF), ou Funções de Base Radial, também pode ser empregada em quase todos os tipos de problemas tratados pela MLP, inclusive aqueles que envolvem aproximação de funções e classificação de padrões (SILVA; SPATTI; FLAUZINO, 2016). A RBF pode ser aplicada em várias áreas, como modelagem, previsão, reconhecimento de padrões etc.; sua estrutura contém três camadas: a primeira é a camada de entrada, que fornece os sinais aplicados externamente, a segunda é a camada oculta, que processa os sinais de entrada, e a última é a camada de saída, que recebe as saídas ponderadas da camada oculta (KUMAR; SRIVASTA; GUPTA, 2018). Trata-se de uma RNA com aprendizagem supervisionada usada para a tarefa de classificação.

- A *Self-Organizing Map*, ou rede SOM, trata-se de uma RNA com aprendizagem não supervisionada (HAYKIN, 2001). As principais aplicações da rede SOM estão no campo de

agrupamento e visualização de dados. Estas características fazem com que a rede SOM seja muito utilizada na tarefa de agrupamento (VESANTO; ALHONIEMI, 2000).

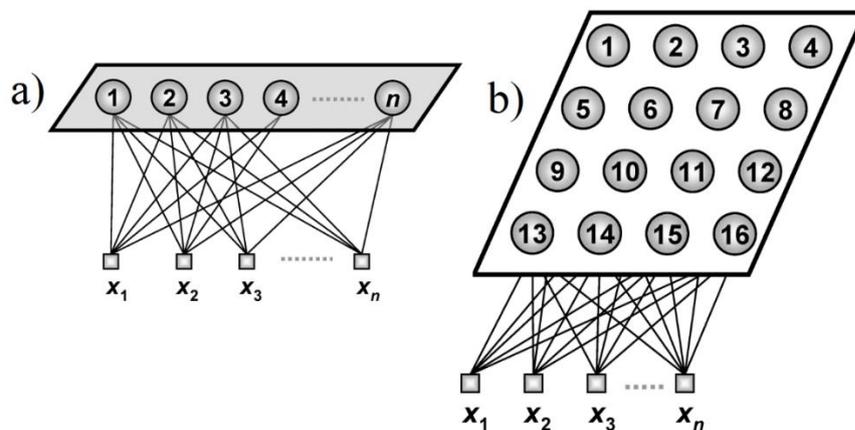
2.6.1.1 Mapa Auto-Organizável de Kohonen (SOM)

O Mapa Auto-Organizável de Kohonen, ou *Self-Organizing Map* (SOM), foi desenvolvido pelo pesquisador finlandês Teuvo Kohonen, no início da década de 1980 (KOHONEN, 1982).

De acordo Vesanto e Alhoniemi (2000), 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 em um conjunto finito de neurônios. Estas características são utilizadas na geração de agrupamentos também chamados de *clusters*. Essa rede tem sido utilizada extensivamente como ferramenta analítica e de visualização em análise exploratória de dados, tendo sido empregada com sucesso em uma profusão de aplicações práticas que vão desde o controle de processos industriais e análises de finanças até o gerenciamento de grandes coleções de documentos (KOHONEN, 2013).

A rede SOM é uma rede neural de aprendizagem competitiva organizada em uma única ou duas camadas (uni ou bidimensional). A primeira camada representa o vetor dos dados de entrada, e a segunda corresponde a uma grade de neurônios, geralmente bidimensional, totalmente conectada aos componentes do vetor de entrada (KOHONEN 2001; CALVO-DAHLBORG; BROWN, 2017). Na Figura 5, mostra-se um exemplo de mapa unidimensional (a) e bidimensional (b).

Figura 5 – Mapa topológico unidimensional e bidimensional



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

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); 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). O BMU é calculado por meio da seguinte Equação (1).

$$\arg \min_j \left(\|w_j - x_i\| \right) \quad (1)$$

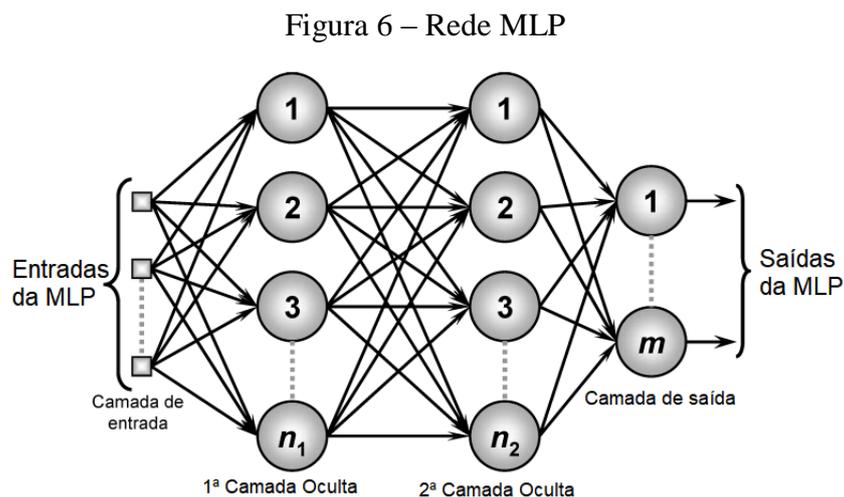
Para cada dado apresentado à rede, haverá uma competição entre todos os neurônios pelo direito de representá-lo. Vence a competição o neurônio que tiver o vetor de pesos mais próximo ao do vetor de dados (HAYKIN, 2001).

As principais aplicações da rede SOM estão no campo 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; ALHONIEMI, 2000).

2.6.1.2 Multilayer Perceptron (MLP)

Uma *Multilayer Perceptron* (MLP) consiste em um conjunto de unidades que constituem a camada de entrada, uma ou mais camadas ocultas, e uma camada de saída, no qual o sinal de entrada se propaga pela RNA camada por camada (HAYKIN, 2001).

Na Figura 6, ilustra-se uma rede MLP com duas camadas ocultas.



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

A MLP possui 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 MLP de modo que ela possa aprender uma determinada função, é chamado algoritmo de treinamento ou de aprendizado (BIGUS, 1996).

O aprendizado de uma MLP 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; SHAW, 2007).

Como resultado do treinamento, a MLP produzirá valores de saída similares aos do conjunto de dados, para valores que sejam iguais às amostras de treinamento. Para valores intermediários, a rede produzirá uma interpolação. Ou seja, a MLP pode aprender por meio de exemplos (SIMÕES; SHAW, 2007).

O algoritmo de retropropagação do erro (*error back-propagation*), utilizado na MLP, consiste, basicamente, em determinar as variações nos pesos sinápticos, a fim de minimizar o erro obtido na saída, por meio do aprendizado do vetor de treinamento (entrada-saída) (RUMELHART; HINTON; WILLIAMS, 1986).

Para isso, o algoritmo se baseia 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 Δw_{ij} proporcional ao gradiente do erro.

O algoritmo de retropropagação do erro funciona da seguinte forma: apresenta-se um padrão à camada de entrada da rede, e esse padrão é processado, camada por camada, até que a camada de saída forneça a resposta processada, f_{MLP} , como mostra a Equação (2):

$$f_{MLP}(x) = \varphi \left(\sum_1^{Non} v_l \cdot \varphi \left(\sum w_{ij} x_l + b_{l0} \right) + b_0 \right) \quad (2)$$

Em que, v_l e w_{ij} são pesos sinápticos; b_{l0} e b_0 são os biases; e φ , a função de ativação, comumente especificada como sendo a função sigmoide.

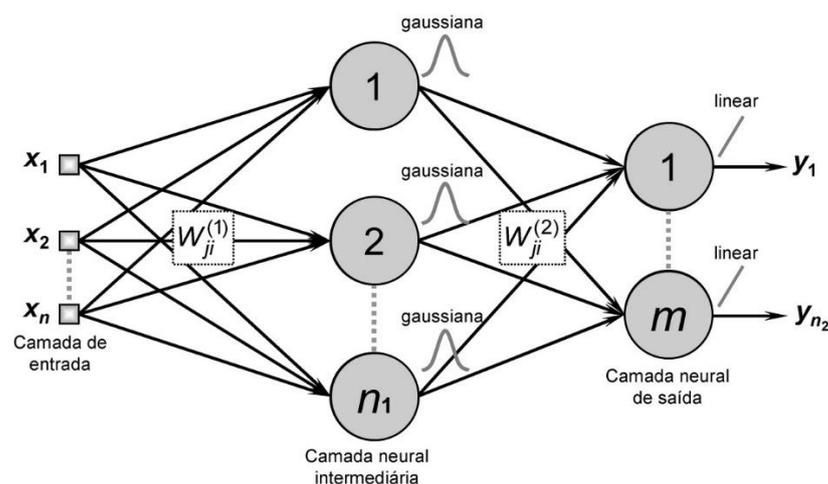
O princípio desse algoritmo, para cálculo dos erros nas camadas intermediárias, é o seu cálculo por retroalimentação, possibilitando, desta forma, o ajuste dos pesos proporcionalmente aos valores das conexões entre camadas. De acordo com Haykin (2001), a MLP possui as seguintes características: função de ativação não linear (sigmoide), uma ou mais camadas de neurônios ocultos e um alto grau de conectividade. Uma MLP treinada com o algoritmo de retropropagação realiza um mapeamento não linear de entrada-saída. Ainda de acordo com Haykin (2001), matematicamente, uma MLP é um método para determinar uma função $f: A \rightarrow B$, em que A é um conjunto de dados, e alguns elementos de B são conhecidos a priori.

2.6.1.3 Radial Basis Function (RBF)

As redes denominadas funções de base radial, convencionalmente conhecidas como RBF (*Radial Basis Function*), podem ser também empregadas em quase todos os tipos de problemas tratados pela MLP, inclusive aqueles que envolvem aproximação de funções e classificação de padrões.

Diferentemente das redes MLP, que podem ser compostas de diversas camadas intermediárias, a estrutura típica de uma rede RBF é composta por apenas uma, na qual as funções de ativação são do tipo gaussiana, conforme pode-se observar na Figura 7.

Figura 7 – Configuração típica de rede RBF



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Uma das principais particularidades da rede RBF está na estratégia de treinamento utilizada para os ajustes dos pesos de suas únicas camadas neurais. Em conformidade ainda ao observado na Figura 7, outro diferencial dessa arquitetura de rede é o tipo de função de ativação assumida nos neurônios da camada intermediária, sendo caracterizada por funções de base radial como as gaussianas.

O princípio de funcionamento das redes RBF é similar àquela da MLP, em que cada uma de suas entradas $\{x_j\}$, representando os sinais advindos da aplicação, será então propagada pela referida camada intermediária, em direção à camada de saída.

Entretanto, diferentemente da MLP, a estratégia de treinamento da RBF é constituída de duas faces ou estágios bem distintos entre si. O primeiro estágio, associado aos ajustes dos pesos dos neurônios da camada intermediária, adota um método de aprendizagem auto-organizado (não-supervisionado), que é dependente apenas das características dos dados de entrada. Na realidade, esse ajuste está diretamente relacionado à alocação das funções de base radial. Já o segundo estágio, vinculado aos ajustes dos pesos dos neurônios da camada de saída,

utiliza um critério de aprendizagem similar àquele usado na última camada da MLP, ou seja, a regra delta generalizada.

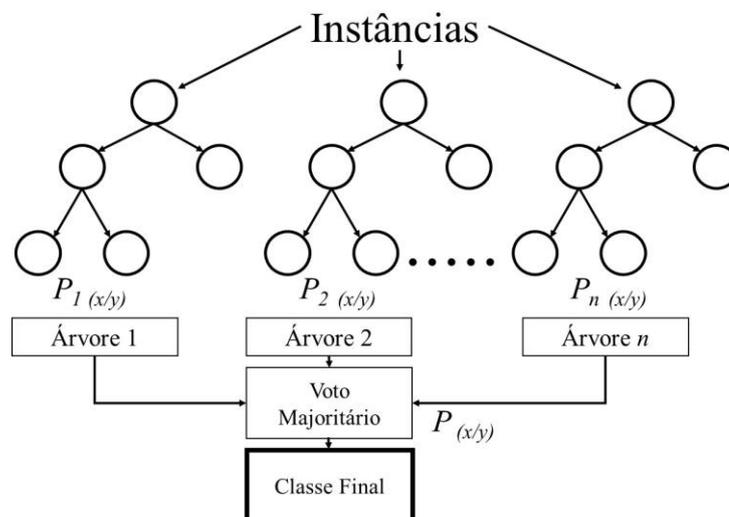
2.6.2 Random Forest (RF)

Random Forest (RF), ou Florestas Aleatórias, é um algoritmo desenvolvido por Breiman (2001), que permite que se obtenha modelos eficazes sem nenhuma preparação de dados ou conhecimento de modelagem.

Uma RF pode ser descrita como um classificador formado por um conjunto de árvores de decisão $\{h(\mathbf{X}, vk), k, 1, \dots\}$, em que vk são vetores aleatórios amostrados de forma independentes, distribuídos igualmente em todas as árvores da floresta. O resultado do processo de classificação é a classe X com maior número de votos dentre todas as árvores consideradas (HAN; KAMBER; PEI, 2012).

A RF é um algoritmo de classificação e regressão não paramétrico e não linear, que gera um conjunto de árvores de decisão. Nos problemas de classificação, o resultado obtido é dado por um sistema de votação, em que a nova classe predita é aquela que obteve o maior número de votos. A RF segue o conceito de aprendizado por conjunto, ou seja, a ideia básica do método é construir um conjunto de árvores de decisão (BREIMAN, 2001). Na Figura 8, ilustra-se, de forma simplificada, esse modelo de aprendizado por conjunto e o sistema de votação.

Figura 8 – Modelo de aprendizado por conjunto da RF



Fonte: Adaptado de Ibrahim e Khatib (2017).

Na árvore de decisão (Árvore 1), demonstrada na Figura 8, o nó de onde a árvore inicia (instâncias) é chamado de nó raiz. O nó em que a cadeia termina é conhecido como nó folha (classe). Duas ou mais ramificações podem ser estendidas de cada nó interno, ou seja, um nó que não é o nó folha. Um nó representa uma certa característica, enquanto as ramificações

representam um intervalo de valores. Esses intervalos de valores atuam como pontos de partição para o conjunto de valores da característica especificada. Cada nó na árvore especifica um teste de algum atributo da instância e cada ramificação descende desse nó e corresponde a um dos valores possíveis para esse atributo. Uma instância é classificada iniciando no nó raiz da árvore, testando o atributo especificado por este nó e movendo-se para baixo no ramo da árvore, correspondente ao valor do atributo no exemplo fornecido. Esse processo é repetido para a subárvore enraizada no novo nó (MITCHELL, 1997).

A RF trabalha em duas bases de aleatoriedade da seguinte forma (BREIMAN, 2001):

a) Cada árvore de decisão é construída individual e aleatoriamente em uma amostra diferente do conjunto de dados de treinamento.

b) Durante todo o processo de construção de cada árvore, um subconjunto de n instâncias é escolhido aleatoriamente no conjunto de dados de treinamento, e o melhor subconjunto de acordo com essas n instâncias é utilizada. Para um novo caso, a previsão da RF é construída pelas árvores de decisão que são agregadas. Na RF que inclui N árvores de decisão, a probabilidade de saída da RF com o rótulo da classe y para um novo caso é calculado conforme a Equação (3), usando o seguinte modelo de aprendizado por conjunto $P(y/x)$.

$$k = \frac{1}{N} \sum_{i=1}^N P_i(y / x) \quad (3)$$

A Equação (3) calcula a média da probabilidade de saída das árvores de decisão geradas a partir de amostras aleatórias do conjunto de dados.

Cada árvore de decisão na RF divide o conjunto de dados em subconjuntos menores, construindo uma subárvore com nós de folha e nós de decisão. Um nó de decisão possui duas ou mais ramificações com folhas. Cada categoria no conjunto de dados é representada pelo nó da folha. Como as árvores de decisão na RF dependem de dados aleatórios, elas podem não ter significado e podem ser ruidosas. Portanto, a RF calcula a média dessas árvores de decisão para construir um modelo de baixa variação. Cada um desses subconjuntos menores é criado por um tipo de amostragem chamado de *bootstrap* (HAN et al., 2011), a qual é o tipo de amostragem com reposição, ou seja, cada novo conjunto poderá ter alguns registros incluídos mais de uma vez, e outros não incluídos nenhuma vez. A amostragem *bootstrap* garante que 1/3 dos exemplos seja usado para testar as árvores após sua construção. As árvores de decisão impróprias podem ser descartadas, e as árvores úteis são usadas para criar o modelo resultante (BREIMAN, 2001).

A partir de cada subconjunto desenvolvido, uma árvore de decisão é criada. A construção das árvores ocorre por meio de uma seleção de atributos aleatória dos subconjuntos, os quais são utilizados nos nós de cada uma das árvores desenvolvidas. Uma floresta aleatória é uma coleção dessas árvores de decisão.

Quando a floresta está formada, há um número grande de árvores de decisão a serem testadas, e todas contribuem para a classificação do objeto em estudo, por meio de um voto sobre qual classe o atributo meta deve pertencer. Cada voto tem um certo “peso”, que é afetado pela similaridade entre cada árvore, sendo que quanto menor a similaridade entre duas árvores melhor, e pela força que cada árvore tem individualmente, ou seja, quanto mais precisa uma árvore for, melhor será sua nota. O ideal é manter a precisão das árvores sem aumentar sua similaridade (HAN et al., 2011).

O algoritmo é escalar e pode lidar com conjuntos com um grande número de atributos. O uso de subconjuntos e amostragem *bootstrap* torna o algoritmo mais robusto do que uma simples árvore, apresentando boa taxa de acerto quando testado em diferentes conjuntos de dados (BREIMAN, 2001).

A RF ainda é computacionalmente muito efetiva, além de evitar sobreajuste (*overfitting*) e ser pouco sensível a ruídos (BREIMAN, 2001).

Abordagens baseadas em RF apresentam uma ampla gama de aplicações, como previsão de demanda de encomendas (ALSANAD, 2018), análise empírica do absentismo no trabalho (PRIYANKA; NAYAK, 2019), avaliação de risco de crédito em empréstimos (PAN; ZHOU, 2019), e previsão de radiação solar (SRIVASTAVA; TIWARI; GIRI, 2019).

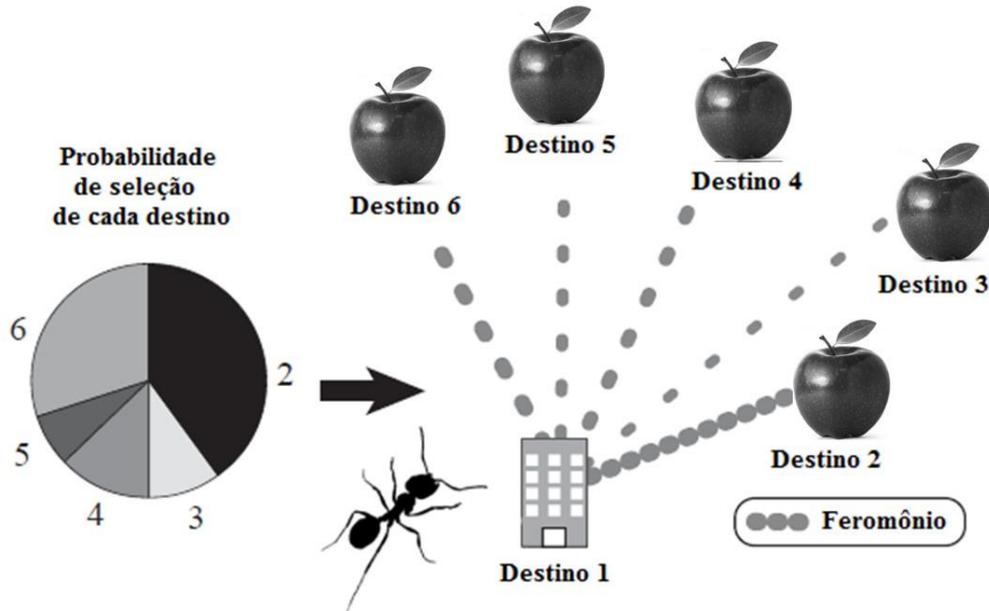
2.6.3 Algoritmo de Otimização por Colônia de Formigas (ACO)

O Algoritmo de Otimização por Colônia de Formigas, também conhecido como *Ant Colony Optimization* (ACO), é uma metaheurística inspirada no comportamento das colônias de formigas na natureza. Muitas colônias de formigas são capazes de cooperar para executar tarefas complexas, apesar da falta de um mecanismo central de controle e apesar de terem um comportamento individual relativamente simples. Um exemplo disso é a existência de espécies de formigas com pouca ou nenhuma visão que são capazes de descobrir o caminho mais curto entre o ninho e uma fonte de alimento (DORIGO; STÜTZLE, 2004).

Toda interação entre formigas individuais e entre o meio ambiente é realizada de maneira indireta, ou seja, as formigas liberam uma substância química chamada feromônio no chão para criar uma trilha. Trilhas com maior concentração de feromônio têm maior probabilidade de serem seguidas pelas demais formigas. De acordo com Yoshikawa e Otani

(2010), os destinos mais curtos com feromônio elevado têm maior probabilidade de seleção. Na Figura 9, ilustra-se a seleção baseada nesse critério.

Figura 9 – Probabilidade de seleção de cada destino



Fonte: Adaptado de Yoshikawa e Otani (2010).

A atratividade de cada destino pode ser descrita conforme a Equação (4):

$$a_{ij}(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i} [\tau_{il}(t)]^\alpha [\eta_{il}]^\beta} \quad (4)$$

A partir da Equação 4, é calculada a atratividade $a_{ij}(t)$ para uma formiga que está em um nó i qualquer, e para todo nó j pertencente a N_i , em que N_i é o conjunto dos nós adjacentes ao nó i . A variável τ_{ij} representa a quantidade de feromônio acumulada no caminho ij , e a variável η_{ij} é igual ao inverso do comprimento do caminho ij . O parâmetro t indica que os valores estão variando a cada iteração (BONABEAU; DORIGO; THERAULAZ, 1999; DORIGO; STÜTZLE, 2004).

Após o cálculo de todos os valores dos destinos incidentes a_i , calcula-se a probabilidade de uma formiga k percorrer o destino ij . Na Equação (5), descreve-se essa probabilidade:

$$p_{ij}^k(t) = \frac{a_{ij}(t)}{\sum_{l \in N_i^k} a_{il}(t)} \quad (5)$$

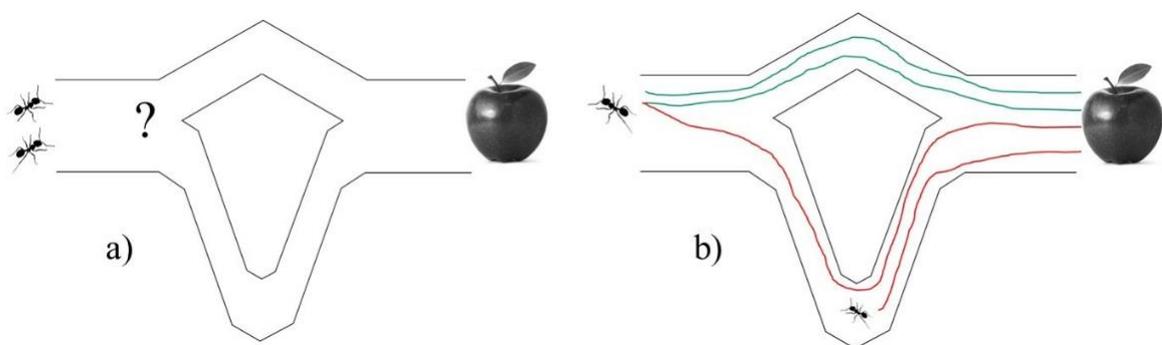
Conforme se pode observar na Equação (5), N_i^k é o subconjunto de N_i , que apresenta apenas os nós que ainda não foram visitados pela formiga k .

A formiga k seleciona um destino entre todas as possíveis alternativas, por meio do cálculo da probabilidade acumulada, que é feito a partir das probabilidades de cada arco P_{ij} ; em seguida, sorteia-se um número aleatório no intervalo entre 0 e 1. O caminho escolhido será aquele em que o número sorteado tenha um valor maior ou igual a P_{ij-1} e menor que P_{ij} . Completadas suas rotas, o feromônio liberado pelas formigas é atualizado em todos os caminhos usados. Esta atualização consiste na evaporação e na deposição de novos feromônios (DORIGO; STÜTZLE, 2004). Na Equação (6) a seguir, descreve-se a atualização dos feromônios:

$$\tau_{ij}(t) \leftarrow (1 - \rho)\tau_{ij}(t) + \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad (6)$$

Como a trilha mais curta é mais rápida, a concentração de feromônios é maior, o que eventualmente resultará na colônia convergindo para usar predominantemente o caminho mais curto (FREITAS; PARPINELLI; LOPES, 2008). Na Figura 10, ilustra-se, de forma simplificada, esse comportamento nas sequências (a) e (b).

Figura 10 – Seleção do caminho das formigas usando o feromônio



Fonte: Adaptado de Dorigo e Stützle (2004) e Martens, Baesens e Fawcett (2011).

Os algoritmos ACO imitam o comportamento das colônias de formigas para realizar uma pesquisa global, que é guiada para buscar regiões do espaço de pesquisa com base na qualidade das soluções. No contexto da tarefa de classificação na mineração de dados, os algoritmos ACO realizam uma pesquisa robusta e flexível para uma boa combinação de

atributos preditores; os algoritmos ACO, para classificação, são criados por meio de regras (OTERO, 2017).

O algoritmo ACO, para classificação, cria regras de classificação *SE-ENTÃO*. A parte *SE* corresponde ao antecedente da regra e contém termos (atributo, valor) que representam testes nos valores dos atributos. A parte *ENTÃO* corresponde ao conseqüente da regra e contém uma previsão de valor da classe. Diz-se que uma instância que satisfaça todos os termos do antecedente de uma regra é coberta pela regra e tem o valor de classe no conseqüente da regra prevista (PARPINELLI; LOPES; FREITAS, 2002). Os algoritmos de classificação do ACO atraíram atenção e um grande número de variações desses algoritmos foi proposto na literatura (MARTENS; BAESENS; FAWCETT, 2011).

Os algoritmos do ACO possuem três características principais:

- a) São algoritmos de base populacional, nos quais as formigas criam soluções para o problema, imitando uma colônia de formigas;
- b) As soluções são criadas por um procedimento probabilístico, em que os componentes da solução são selecionados com base em valores de informações de feromônio e heurística;
- c) Os valores de feromônios são atualizados a cada iteração, usando a qualidade das soluções candidatas (OTERO, 2017).

Em Dorigo e Stützle (2004), são apresentadas outras aplicações dos algoritmos ACO, listados de acordo com o tipo de problema abordado. Abordagens baseadas em ACO já apresentam uma gama de aplicações em mineração de dados, como previsão de crise financeira (UTHAYAKUMAR et al., 2020), previsão da rotatividade de clientes do setor de telecomunicações (ÖZMEN et al., 2019), previsão de inundações fluviais (SABRI; SAIAN, 2017), e tarefas de classificação em mineração de dados (SAMEH, A.; MAGDY, 2010; OTERO; FREITAS, 2013; OTERO; FREITAS; JOHNSON, 2013; OTERO, 2017; AL-BEHADILI; KU-MAHAMUD; SAGBAN, 2018).

3. MATERIAIS E MÉTODOS

Neste capítulo, apresentam-se os materiais e métodos utilizados para desenvolvimento do estudo, a descrição dos experimentos computacionais e a descrição das bases de dados e dos softwares.

3.1 CARACTERIZAÇÃO METODOLÓGICA

Foi realizada uma pesquisa bibliográfica elaborada a partir de artigos de periódicos, livros, teses, dissertações e materiais disponibilizados em bases de dados com conteúdo sobre absenteísmo, presenteísmo, antropometria, ergonomia, mineração de dados e técnicas inteligentes (GIL, 2018).

A pesquisa foi caracterizada como experimental, uma vez que consiste em determinar um objeto de estudo, selecionar as variáveis capazes de influenciá-lo e definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Trata-se de um tipo de pesquisa que procura revelar as relações de causa e efeito entre variáveis (APPOLINÁRIO, 2011; GIL, 2018). Pode também ser caracterizada como exploratória, pois para atingir o objetivo deste trabalho foi preciso analisar o estado da arte e levantamento bibliográfico (GIL, 2018).

Quanto à sua natureza, a pesquisa foi classificada como aplicada, visto que tem como objetivo gerar conhecimento para a resolução de problemas ou necessidades concretas e imediatas, permitindo uma aplicação prática. Muitas vezes, nessa modalidade de pesquisa, os problemas emergem do contexto profissional e podem ser sugeridos pela instituição para que o pesquisador solucione uma situação-problema (YIN, 2015).

Quanto à sua abordagem, a pesquisa foi classificada como quantitativa. De acordo com Rodrigues e Limena (2006), a pesquisa quantitativa é compreendida quando a abordagem está relacionada à quantificação, análise e interpretação de dados obtidos mediante pesquisa voltada para análise e interpretação dos resultados.

3.2 PESQUISA BIBLIOGRÁFICA

A pesquisa bibliográfica foi realizada com as seguintes palavras-chave: Antropometria, Ergonomia, Absenteísmo, Presenteísmo, Mineração de Dados, Redes Neurais Artificiais, Aprendizado de Máquina e Inteligência Artificial. A pesquisa também foi feita na

língua inglesa com as seguintes palavras-chave: *Anthropometry, Ergonomics, Absenteeism, Presenteeism, Data Mining, Artificial Neural Networks, Machine Learning* e *Artificial Intelligence*.

Mediante essa busca, foram encontrados artigos na língua portuguesa com as seguintes palavras combinadas: absenteísmo, presenteísmo e mineração de dados. Já na língua inglesa, foram encontrados artigos com as seguintes palavras combinadas: *Absenteeism, Presenteeism, Data Mining, Artificial Neural Networks, Machine Learning* e *Artificial intelligence*. Vale ressaltar que não foram encontrados trabalhos com todos os termos pesquisados conjuntamente.

A obtenção dos artigos deu-se nas seguintes bases de dados: Science Direct, SciELO, IEEE Xplore e Google Acadêmico. Em relação às palavras-chave Antropometria e Ergonomia, vale ressaltar que não houve um resultado dessas palavras em conjunto com os demais termos utilizados na pesquisa bibliográfica.

Com a pesquisa, 80 artigos foram compilados, mas apenas 30 foram selecionados, após a aplicação dos critérios de inclusão, descritos a seguir:

Critério de inclusão

- Artigos escritos somente na língua inglesa e portuguesa;
- Artigos relacionados somente com os seguintes aspectos comuns:
 - a) Absenteísmo relacionado à saúde no trabalho;
 - b) Absenteísmo relacionado à segurança do trabalho;
 - c) Absenteísmo relacionado à saúde pública;
- Artigos publicados somente a partir de 2012 até janeiro de 2020.

A limitação aos últimos nove anos foi escolhida a partir da primeira publicação em 2012, que contempla absenteísmo, rede neural artificial e lógica *fuzzy*, sendo esse o artigo seminal desta pesquisa Martiniano et al. (2012), até janeiro de 2020, para que o estudo apresentasse o estado da arte.

Na Tabela 1, são apresentados os 30 artigos selecionados na pesquisa bibliográfica, considerando os critérios de exclusão. Em negrito, destacam-se os artigos que mais se aproximam deste trabalho, uma vez que aplicam mineração de dados, aprendizado de máquina e técnicas de inteligência artificial no absenteísmo ou presenteísmo.

Tabela 1 – Relação dos 30 artigos selecionados na pesquisa bibliográfica

ANO	TÍTULO	PUBLICAÇÃO	AUTOR (ES)
2020	Predictive Analysis of Absenteeism in MNCS Using Machine Learning Algorithm	Proceedings of ICRIC 2019. Springer, Cham, 2020	Tewari, Vandita e Jain (2020)
2020	Absenteeism Prediction: A Comparative Study Using Machine Learning Models	International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions	Dogruyol e Sekeroglu (2020)
2019	Employees Absenteeism Factors Based on Data Analysis and Classification	Special Issue in Communication and Information Technology	Asiri, e Abdullah (2019)
2019	Anticipating Next Public Administration Employees Absence Duration	Public administration issues	Costa, Ramos e Moro (2019)
2019	Absenteeism prediction in call center using machine learning algorithms	Conference on Information Systems and Technologies	De Oliveira et al. (2019)
2019	Performance Analysis and Evaluation of Clustering Algorithms	International Journal of Innovative Technology and Exploring Engineering (IJITEE)	Patel e Patel (2019)
2019	Predicting Absenteeism at Work Using Tree-Based Learners	International Conference on Machine Learning and Soft Computing	Wahid et al. (2019)
2019	A hybrid approach of intelligent systems to help predict absenteeism at work in companies	SN Applied Sciences	Araujo et al. (2019)
2019	Empirical Analysis of Absenteeism at Work Place Using Machine Learning	International Conference on Application of Robotics in Industry using Advanced Mechanisms	Priyanka e Nayak (2019)
2019	When do bundles of high performance work systems reduce employee absenteeism? The moderating role of workload	The International Journal of Human Resource Management	De Reuver, Van de Voorde e Kilroy (2019)
2019	Loss of work productivity in a warming world: Differences between developed and developing countries	Jornal of Cleaner Production	Shuang et al. (2019)
2018	Integrating Public Service Motivation in the Job-Demands-Resources Model: An Empirical Analysis to Explain Employees' Performance, Absenteeism, and Presenteeism	International Public Management Journal	Gross, Thaler e Winter (2018)
2018	Perceived workplace adversity and correctional officer psychological well-being: an international examination of the impact of officer response styles.	The Journal of Forensic Psychiatry & Psychology	Trounson, Pfeife e Skues (2018)
2018	Productivity loss of caregivers of schizophrenia patients: a cross-sectional survey in Japan	Journal of Mental Health	Sruamsiri, Mori e Mahlic (2018)
2018	Exploring the Optimal Design of an Employer-Sponsored Sickness-Disability Compensation Insurance Plan When Sickness Presenteeism Is Penalized	North American Actuarial Journal	Ramsay, Oguledo e Krutto (2018)

Fonte: Elaborado pelo autor (2020).

Tabela 1 – Relação dos 30 artigos selecionados na pesquisa bibliográfica (Continuação)

ANO	TÍTULO	PUBLICAÇÃO	AUTOR (ES)
2018	Evaluation of Clustering Algorithms on Absenteeism at Work Dataset	International Journal for Scientific Research & Development	Paakaala, Macha e Mudigonda (2018)
2018	A utilização do algoritmo de árvore de decisão C4.5 para a gestão do absentéismo organizacional	Postal Brasil – Revista técnico científica dos Correios	Andrade (2018)
2018	Artificial neural network and their application in the prediction of Absenteeism at work	International Journal of Recent Scientific Research Research	Ferreira et al. (2018)
2018	Data mining of Absentee data to increase productivity	International Journal of Engineering and Techniques	Gayathri (2018)
2018	Relationships between presenteeism and work-related musculoskeletal disorders among physical therapists in the Republic of Korea	International Journal of Occupational Safety and Ergonomics	Bae (2018)
2018	Sickness presenteeism: measurement and management challenges	Ergonomics	Whysall, Bowden e Hewitt (2018)
2018	Absentéismo por distúrbios musculoesqueléticos em trabalhadores do Brasil: milhares de dias de trabalho perdidos	Revista Brasileira de Epidemiologia	Haeffner et al. (2018)
2018	A Multi-Level Integration of the Antecedents and Dynamics of Absenteeism	Academy of Management	Miraglia e Johns (2018)
2018	Predictors of disability and absenteeism in workers with non-specific low back pain: A longitudinal 15-month study	Applied Ergonomics	Lardon et al. (2018)
2017	Uma Abordagem Fuzzy para o estudo do Presenteísmo.	Revista Pensamento Contemporâneo em Administração	Lopes et al. (2017)
2017	Sickness presenteeism and sickness absence over time: a UK employee perspective	Work & Stress	Collins, Cartwright e Cowlishaw (2017)
2016	Um estudo usando data mining com o algoritmo density based scan algorithm with noise (DBSCAN) na clusterização de perfis de empregados absentéistas e presenteístas	13th CONTECSI - International Conference on Information Systems and Technology Management	Ferreira, Martiniano e Sassi (2016a)
2016	Dashboard Inteligente para Apoio à Tomada de Decisão em Empresa de Courier	Revista Gestão & Tecnologia	Ferreira, Martiniano e Sassi (2016b)
2014	Employees' attendance patterns prediction using classification algorithm case study: a private company in Indonesia	Int'l Journal of Computing, Communications & Instrumentation Engg (IJCCIE)	Qomariyah e Suchyo (2014)
2012	Application of a neuro fuzzy network in prediction of absenteeism at work	7ª Conferencia Ibérica de Sistemas y Tecnologías de Información - CISTI	Martiniano et al. (2012)

Fonte: Elaborado pelo autor (2020).

Dos 30 artigos selecionados, somente 18, que estão destacados em negrito utilizam técnica inteligente aplicada na área de gestão de pessoas. Além disso, somente 12, dos artigos compilados, utilizaram a base de dados denominada *Absenteeism at Work Data set* (<http://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>), cedida em 2018 para o repositório de dados de *Machine Learning* da Universidade da Califórnia Irvine (*UCI Machine Learning Repository*). A base de dados é oriunda do seguinte artigo: “Martiniano, A., Ferreira, R. P., Sassi, R. J.; Affonso, C. *Application of a neuro fuzzy network in prediction of absenteeism at work*. In Information Systems and Technologies IEEE (CISTI), 7th Iberian Conference. pp. 1-4, 2012”, no qual a autora deste trabalho também é coautora. Informações adicionais sobre a base de dados cedida podem ser encontradas na seção 5.2.

No Quadro 3, apresenta-se a relação entre autor (es), palavras-chave, técnicas aplicadas, idioma base de dados dos 30 artigos selecionados.

Quadro 3 – Relação entre autor (es), palavras-chave, técnicas aplicadas, idioma e base de dados dos artigos selecionados

Ano Autor (es)		Palavras-chave					Técnicas aplicadas					Idioma	Base de dados	
		Absenteísmo	Presenteísmo	Mineração de dados	Redes Neurais Artificiais	Aprendizado de máquina	Inteligência Artificial	Lógica Fuzzy	Random Forest	Árvore de decisão	Support Vector Machines			DBSCAN
2020	Tewari, Vandita e Jain (2020)	X		X		X				X			I	U
2020	Dogruyol e Sekeroglu (2020)	X			X								I	U
2019	Asiri, e Abdullah (2019)	X				X		X	X				I	U
2019	Costa, Ramos e Moro (2019)	X		X						X			I	O
2019	De Oliveira et al. (2019)	X			X	X		X		X			I	O
2019	Patel e Patel (2019)	X		X		X	X		X				I	U
2019	Wahid et al. (2019)	X				X		X	X				I	U
2019	Araujo et al. (2019)	X			X		X						I	U
2019	Priyanka e Nayak (2019)	X	X		X	X		X	X	X			I	U
2019	De Reuver, Van de Voorde e Kilroy (2019)	X											I	O
2019	Shuang et al. (2019)	X	X										I	O
2018	Gross, Thaler e Winter (2018)	X	X										I	O
2018	Trounson, Pfeife e Skues (2018)	X	X										I	O
2018	Sruamsiri, Mori e Mahlic (2018)	X	X										I	O
2018	Ramsay, Oguledo e Krutto (2018)	X	X										I	O

Fonte: Elaborado pelo autor (2020).

Quadro 3 – Relação entre autor (es), palavras-chave, técnicas aplicadas, idioma e base de dados dos artigos selecionados (Continuação)

Ano Autor (es)		Palavras-chave						Técnicas aplicadas						Idioma	Base de dados
		Absenteísmo	Presenteísmo	Mineração de dados	Redes Neurais Artificiais	Aprendizado de máquina	Inteligência Artificial	Lógica Fuzzy	Random Forest	Árvore de decisão	Support Vector Machines	DBSCAN	Rough Sets	Português (P) Inglês (I)	UCI (U) Outras bases (O)
2018	Paakaala, Macha e Mudigonda (2018)	X		X		X					X		I	U	
2018	Andrade (2018)	X		X					X				P	O	
2018	Ferreira et al. (2018)	X			X							X	I	O	
2018	Gayathri (2018)	X		X	X	X			X				I	U	
2018	Bae (2018)	X	X										I	O	
2018	Whysall, Bowden e Hewitt (2018)	X	X										I	O	
2018	Haefner et al. (2018)	X											P	O	
2018	Miraglia e Johns (2018)	X	X										I	O	
2018	Lardon et al. (2018)	X											I	O	
2017	Lopes et al. (2017)	X	X					X					P	O	
2017	Collins, Cartwright e Cowlshaw (2017)	X	X										I	O	
2016	Ferreira, Martiniano e Sassi (2016a)	X	X	X							X		P	U	
2016	Ferreira, Martiniano e Sassi (2016b)	X			X		X						P	U	
2014	Qomariyah e Suchahyo (2014)	X		X					X				I	O	
2012	Martiniano et al. (2012)	X			X		X	X					P	U	

Fonte: Elaborado pelo autor (2020).

Após a realização da pesquisa bibliográfica, constatou-se a inexistência de estudos envolvendo a associação de dados antropométricos, ergonômicos, absenteístas e presenteístas. Também foi possível comprovar a falta de estudos aplicando qualquer técnica de mineração de

dados em bases de dados com esses dados associados e, tampouco, no auxílio à previsão do comportamento dos empregados no trabalho.

3.3 BASE DE DADOS E PLATAFORMA DE EXPERIMENTOS

A plataforma de hardware utilizada nos experimentos foi um microcomputador com processador Intel® Core™ CPU de 2.00 GHz, com 4 GB de memória RAM e 1 TB de disco rígido, sistema operacional de 64 bits.

No Quadro 4, apresentam-se os softwares utilizados neste estudo, considerando utilização, descrição e website.

Quadro 4 – Softwares utilizados no trabalho

Software	Utilização	Descrição	Website
MBP	Utilizado nos experimentos computacionais para a aplicação da MLP 1 e MLP 2 no enriquecimento dos dados.	O <i>Multiple Back-Propagation</i> é um software livre para desenvolvimento de experimentos computacionais com redes neurais artificiais (MBP, 2020).	http://mbp.sourceforge.net/
SOMine	Utilizado nos experimentos computacionais para aplicação da rede SOM no enriquecimento dos dados.	Trata-se de um software pago para realização de experimentos com a rede SOM (KOHONEN, 2001; VISCOVERY, 2020). Existe uma versão livre, porém com recursos limitados à disposição.	https://www.viscovery.net/
WEKA	Utilizado nos experimentos computacionais para aplicação da <i>Random Forest</i> e das Redes Neurais Artificiais do tipo MLP e RBF no auxílio à previsão do comportamento no trabalho. Também foi utilizado para gerar as matrizes de confusão e o índice Kappa.	<i>Knowledge Analysis</i> ou WEKA (WEKA, 2020) é um software livre que oferece implementações de diversos algoritmos para mineração de dados. Fornece as funcionalidades para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização. (NAIK e SAMANT, 2016). Em Witten et al., (2016), são apresentadas as funcionalidades do software WEKA com detalhes.	https://www.cs.waikato.ac.nz/ml/weka/
MYRA	Utilizado nos experimentos computacionais para aplicação do ACO para auxiliar na previsão do comportamento no trabalho.	É um software livre para desenvolvimento de experimentos computacionais com o ACO em mineração de dados na tarefa de classificação por meio de regras (OTERO, 2017).	https://github.com/febo/myra

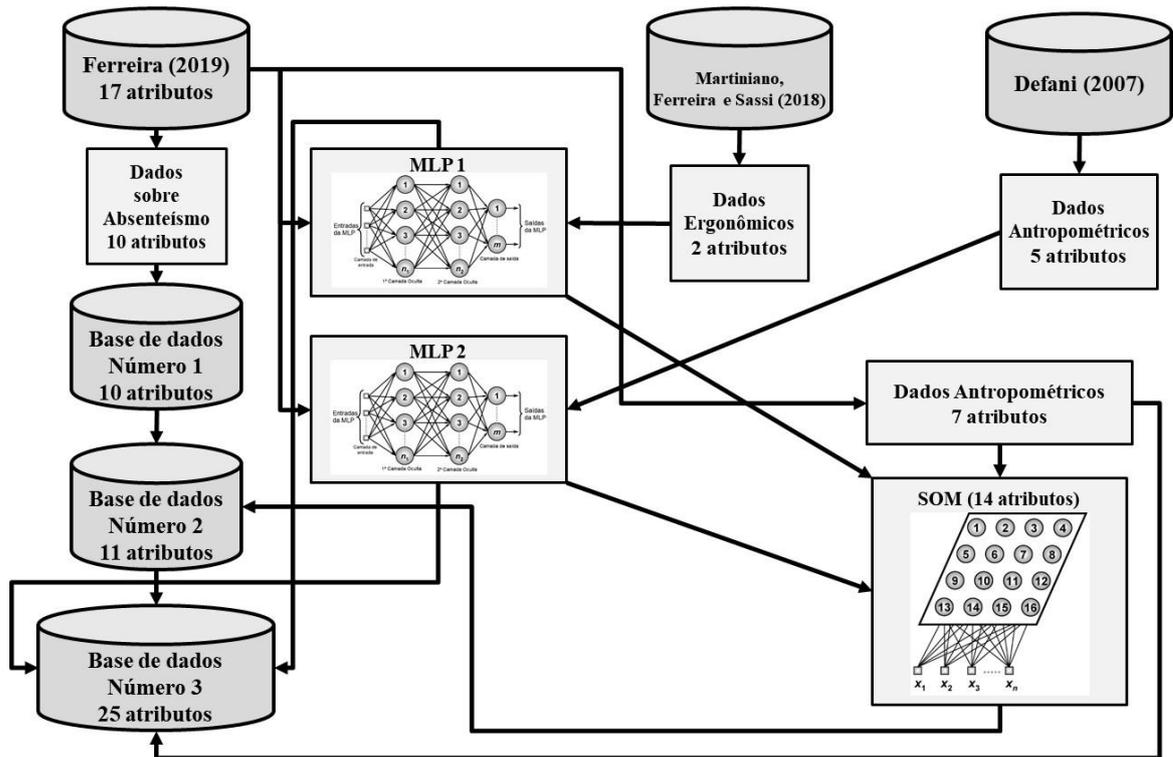
Fonte: Elaborado pelo autor (2020).

Os experimentos computacionais, desenvolvidos para auxiliar na previsão do comportamento dos empregados, foram realizados em três bases de dados distintas:

- a base de dados número 1 possui 10 atributos e 2.403 registros;
- a base de dados número 2 possui 11 atributos e 2.403 registros;
- a base de dados número 3 possui 25 atributos e 2.403 registros.

Para chegar até essas três bases de dados, foram realizados os procedimentos descritos a seguir na Figura 11.

Figura 11 – Procedimentos para o enriquecimento da base de dados



Fonte: Elaborado pelo autor (2020).

As três bases de dados são apresentadas a seguir:

- a) Base de dados oriunda da tese de doutorado de Ferreira (2019): A base de dados foi escolhida por se tratar de uma base composta por atributos relacionados ao absenteísmo e presenteísmo, além de possuir atributos relacionados à antropometria. A base de dados da tese possui 2.403 registros e 50 atributos coletados durante 10 anos, no período de janeiro de 2008 a dezembro de 2017. Neste trabalho, foram utilizados 17 atributos e os 2.403 registros. A base de dados já tinha sido reduzida por meio de *Rough Sets* (RS). Dessa forma, os atributos de maior relevância já eram conhecidos. Por esse motivo, optou-se por não utilizar a base de dados inteira com os 50 atributos.
- b) Base de dados oriunda da dissertação de mestrado de Defani (2007): A base de dados foi escolhida por se tratar de uma base composta por atributos relacionados a medidas antropométricas de empregados brasileiros, ou seja, por assemelhar-se ao perfil antropométrico dos empregados da base de dados de Ferreira (2019). A base de dados dessa dissertação possui 640 registros e 6 atributos. Todos os registros e atributos foram utilizados.

- c) Base de dados oriunda do artigo científico de Martiniano, Ferreira e Sassi (2018): A base de dados foi escolhida por se tratar de uma base composta por atributos relacionados à ergonomia e por pertencer a mesma empresa estudada por Ferreira (2019), ou seja, os empregados executam as mesmas atividades laborais e passaram pelos mesmos treinamentos corporativos relacionados à postura ergonômica. A base de dados desse artigo possui 132 registros e 4 atributos. Todos os registros e atributos foram utilizados.

Essas três bases de dados são de fonte secundária. Sendo assim, não foi necessária submetê-las a um comitê de ética em pesquisa (SILVA; CAVEIÃO, 2016). Segundo Mattar (2014), dados secundários são aqueles que já foram coletados, tabulados, ordenados e até analisados, ou seja, já existem e já foram publicados com propósitos de atender às necessidades de uma pesquisa.

No Quadro 5, apresentam-se a descrição e a origem dos atributos por base de dados.

Quadro 5 – Descrição e origem dos atributos por base de dados

Item	Descrição dos atributos	Origem do atributo por base de dados
1	Corresponde ao código individual	Ferreira (2019)
2	Corresponde aos motivos do absenteísmo (formado pela união dos atributos da Tabela 2 e da Tabela 3)	Ferreira (2019)
3	Corresponde ao tempo de permanência do indivíduo na base de dados em anos	Ferreira (2019)
4	Corresponde ao dia da semana	Ferreira (2019)
5	Corresponde ao tempo de afastamento em dias e horas, medido e transformado em horas	Ferreira (2019)
6	Corresponde à quantidade de licenças médicas registradas por mês	Ferreira (2019)
7	Corresponde ao tempo de serviço em anos completos	Martiniano, Ferreira e Sassi (2018) e Ferreira (2019)
8	Corresponde à idade em anos completos	Martiniano, Ferreira e Sassi (2018) e Ferreira (2019)
9	Corresponde à frequência das licenças médicas por indivíduo	Ferreira (2019)
10	Corresponde à classificação do indivíduo (presenteísta, normal e absenteísta)	Ferreira (2019)
11	Corresponde ao peso em quilograma	Ferreira (2019)
12	Corresponde à estatura do empregado em centímetros (altura)	Defani (2007) e Ferreira (2019)
13	Corresponde ao comprimento do braço	Defani (2007)
14	Corresponde à altura dos ombros (do pé até o ombro)	Defani (2007)
15	Corresponde à altura do cotovelo (do pé até o cotovelo)	Defani (2007)
16	Corresponde à altura do púbis (do pé até a região pubiana)	Defani (2007)
17	Corresponde à altura do epigástrico (do pé até a altura da região epigástrica)	Defani (2007)
18	Corresponde ao IMC – Valores contínuos	Ferreira (2019)
19	Corresponde ao IMC – Valores categóricos	Ferreira (2019)
20	Corresponde ao tamanho do pé	Ferreira (2019)
21	Corresponde ao tamanho do tronco	Ferreira (2019)
22	Corresponde ao tamanho da cintura	Ferreira (2019)
23	Corresponde ao conhecimento sobre ergonomia	Martiniano, Ferreira e Sassi (2018)
24	Corresponde à prática de ergonomia no dia a dia	Martiniano, Ferreira e Sassi (2018)
25	Características antropométricas e ergonômicas – Atributo enriquecido por meio da MLP 1, MLP 2 e rede SOM	Autor

Fonte: Elaborado pelo autor (2020).

Os atributos 7 e 8, destacados em negrito, são comuns às duas bases de dados de Martiniano, Ferreira e Sassi (2018) e Ferreira (2019), assim como o Item 12, também em negrito, é comum a duas bases de dados de Defani (2007) e Ferreira (2019).

O atributo listado no Item 25, características antropométricas e ergonômicas, foi formado pelo enriquecimento por meio da MLP 1, MLP 2 e rede SOM, para compor a base de dados a fim de auxiliar na previsão do comportamento no trabalho.

O segundo atributo (Item 2), apresentado no Quadro 5, refere-se aos motivos do absenteísmo estratificados em vinte e sete categorias, sendo vinte e uma com Classificação Internacional de Doenças (CID) (Tabela 2) e seis sem CID (Tabela 3). Juntas, essas categorias formam o atributo (Item) 2.

Tabela 2 – Classificação Internacional de Doenças (CID)

Capítulo	Descrição
I	Algumas doenças infecciosas e parasitárias (A00-B99)
II	Neoplasias [tumores] (C00-D48)
III	Doenças do sangue e dos órgãos hematopoiéticos e alguns transtornos imunitários (D50-D89)
IV	Doenças endócrinas, nutricionais e metabólicas (E00-E90)
V	Transtornos mentais e comportamentais (F00-F99)
VI	Doenças do sistema nervoso (G00-G99)
VII	Doenças do olho e anexos (H00-H59)
VIII	Doenças do ouvido e da apófise mastoide (H60-H95)
IX	Doenças do aparelho circulatório (I00-I99)
X	Doenças do aparelho respiratório (J00-J99)
XI	Doenças do aparelho digestivo (K00-K93)
XII	Doenças da pele e do tecido subcutâneo (L00-L99)
XIII	Doenças do sistema osteomuscular e do tecido conjuntivo (M00-M99)
XIV	Doenças do aparelho geniturinário (N00-N99)
XV	Gravidez, parto e puerpério (O00-O99)
XVI	Algumas afecções originadas no período perinatal (P00-P96)
XVII	Malformações congênitas, deformidades e anomalias cromossômicas (Q00-Q99)
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte (R00-R99)
XIX	Lesões, envenenamento e algumas outras consequências de causas externas (S00-T98)
XX	Causas externas de morbidade e de mortalidade (V01-Y98)
XXI	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde (Z00-Z99)

Fonte: Adaptado de WHO (2019).

A Classificação Internacional de Doenças e Problemas Relacionados à Saúde, também conhecida como Classificação Internacional de Doenças (CID), é publicada pela Organização Mundial de Saúde (OMS), do inglês *World Health Organization* (WHO), e visa padronizar a codificação de doenças e outros problemas relativos à saúde.

A CID fornece códigos relativos à classificação de doenças e a uma grande variedade de sinais, sintomas, aspectos anormais, queixas, circunstâncias sociais e causas externas para

ferimentos ou doenças. A cada estado de saúde é atribuída uma categoria única à qual corresponde um código ou uma classificação CID (WHO, 2019).

Na Tabela 3, apresentam-se os seis motivos (categorias) de absenteísmo não classificados pela CID.

Tabela 3 – Lista dos seis motivos (categorias) sem CID

Motivo/Categoria	Descrição
I	Acompanhante
II	Consulta médica
III	Doador de sangue
IV	Exame
V	Fisioterapia
VI	Consulta odontológica

Fonte: Elaborado pelo autor (2020).

Os motivos do absenteísmo sem CID foram incluídos na base de dados pelo fato de que outros motivos do absenteísmo, além dos classificados pela CID, também são relevantes para a composição do trabalho.

As categorias de absenteísmo não classificadas pela CID são descritas a seguir:

- Acompanhante: Absenteísmo relacionado ao acompanhamento de familiares, como genitores idosos ou progenitores, cônjuge ou filhos menores de idade, com a devida comprovação. Normalmente previsto em acordo coletivo de trabalho.

- Consulta médica: Absenteísmo de curta duração (menor que um dia ou oito horas) para realização de consultas médicas que não apresentam a notação de CID;

- Doador de sangue: O empregado poderá deixar de comparecer ao serviço sem prejuízo de salário, por um dia, a cada 12 (doze) meses de trabalho, em caso de doação voluntária de sangue, devidamente comprovada. Trata-se de absenteísmo amparado pela LEI 1075/1950, Inciso IV (SANTOS, 2015);

- Exame: Absenteísmo de curta duração para realização de exames diversos sem a notação de CID;

- Fisioterapia: Absenteísmo de curta duração para realização de fisioterapia;

- Consulta odontológica: Absenteísmo de curta duração para realização de tratamento odontológico.

3.3.1 Características dos atributos das bases de dados selecionadas

No Quadro 6, são apresentadas a descrição e as características dos atributos das bases de dados selecionadas, considerando o tipo de variável, o intervalo, a origem dos atributos.

Quadro 6 – Características dos atributos das bases de dados

Item	Descrição dos atributos	Tipo de variável	Intervalo	Origem do atributo por base de dados (Treinamento)	Origem do atributo por base de dados (Teste)
1	Corresponde ao código individual	Catégorica/Discreta	{1...39}	Ferreira (2019)	Ferreira (2019)
2	Corresponde aos motivos do absentéismo	Catégorica/Discreta	{1...27}	Ferreira (2019)	Ferreira (2019)
3	Corresponde ao tempo de permanência do indivíduo na base de dados em anos	Númérica/Contínua	{1,6...10}	Ferreira (2019)	Ferreira (2019)
4	Corresponde ao dia da semana	Catégorica/Discreta	{2...6}	Ferreira (2019)	Ferreira (2019)
5	Corresponde ao tempo de afastamento em dias e horas medido e transformado em horas	Númérica/Discreta	{1...120}	Ferreira (2019)	Ferreira (2019)
6	Corresponde à quantidade de licenças médicas registradas por mês	Númérica/Discreta	{4...48}	Ferreira (2019)	Ferreira (2019)
7	Corresponde ao tempo de serviço em anos completos	Númérica/Discreta	{0...57}	Martiniano, Ferreira e Sassi (2018)	Ferreira (2019)
8	Corresponde à idade em anos completos	Númérica/Discreta	{24...73}	Martiniano, Ferreira e Sassi (2018)	Ferreira (2019)
9	Corresponde à frequência das licenças médicas por indivíduo	Númérica/Discreta	{0...1196}	Ferreira (2019)	Ferreira (2019)
10	Corresponde à classificação do indivíduo (presenteísta, normal e absentéista)	Catégorica/Discreta	{1...3}	Ferreira (2019)	Ferreira (2019)
11	Corresponde ao peso em quilograma	Númérica/Discreta	{56...110}	Ferreira (2019)	Ferreira (2019)
12	Corresponde à estatura do empregado em centímetros (altura/entrada)	Númérica/Contínua	{1,63...1,96}	Defani (2007)	Ferreira (2019)
13	Corresponde ao comprimento do braço	Númérica/Contínua	{0,69...0,81}	Defani (2007)	Ferreira (2019)
14	Corresponde à altura dos ombros (do pé até o ombro)	Númérica/Contínua	{1,34...1,53}	Defani (2007)	Ferreira (2019)
15	Corresponde à altura do cotovelo (do pé até o cotovelo)	Númérica/Contínua	{1,00...1,14}	Defani (2007)	Ferreira (2019)
16	Corresponde à altura do púbis (do pé até a região pubiana)	Númérica/Contínua	{0,83...0,93}	Defani (2007)	Ferreira (2019)
17	Corresponde à altura do epigástrico (do pé até a altura da região epigástrica)	Númérica/Contínua	{1,11...1,28}	Defani (2007)	Ferreira (2019)
18	Corresponde ao IMC – Valores contínuos	Númérica/Contínua	{19,2...38}	Ferreira (2019)	Ferreira (2019)
19	Corresponde ao IMC – Valores catégoricos	Catégorica/Discreta	{1...4}	Ferreira (2019)	Ferreira (2019)
20	Corresponde ao tamanho do pé	Númérica/Discreta	{38...43}	Ferreira (2019)	Ferreira (2019)
21	Corresponde ao tamanho do tronco	Númérica/Discreta	{3...6}	Ferreira (2019)	Ferreira (2019)
22	Corresponde ao tamanho da cintura	Númérica/Discreta	{38...58}	Ferreira (2019)	Ferreira (2019)
23	Corresponde ao conhecimento sobre ergonomia	Númérica/Discreta	{1...10}	Martiniano, Ferreira e Sassi (2018)	Ferreira (2019)
24	Corresponde à prática de ergonomia no dia a dia	Númérica/Discreta	{1...10}	Martiniano, Ferreira e Sassi (2018)	Ferreira (2019)
25	Características antropométricas e ergonômicas (Rede SOM)	Catégorica/Discreta	{1...3}	Autor	Autor

Fonte: Elaborado pelo autor (2020).

3.3.2 Descrição do enriquecimento das bases de dados

A base de dados enriquecida por meio da aplicação da rede SOM foi formada pelos atributos relacionados às medidas antropométricas das bases de dados de Ferreira (2019) e Defani (2007), somados aos atributos relacionados ao conhecimento e à prática sobre ergonomia da base de dados de Martiniano, Ferreira e Sassi (2018). Juntos, os atributos chegaram a um total de 14, servindo assim para o enriquecimento da base de dados. A rede SOM gerou três agrupamentos que formaram um único atributo: Características antropométricas e ergonômicas (Item 25).

- a) A base de dados enriquecida por meio da aplicação da MLP 1 foi formada pelos atributos da base de dados de Martiniano, Ferreira e Sassi (2018) e Ferreira (2019). Com isso, chegou-se a um total de 4 atributos, que serviram para o enriquecimento, por meio da inferência do conhecimento e prática sobre ergonomia, da base de dados de Ferreira (2019), uma vez que as bases de dados possuíam os atributos Tempo de serviço em anos e Idade em anos como atributo comum (Itens 7 e 8).
- b) A base de dados enriquecida por meio da rede MLP 2 foi formada pelos atributos da base de dados de Defani (2007) e Ferreira (2019). Juntos, os atributos chegaram a um total de 5 e serviram para o enriquecimento, por meio da inferência das medidas antropométricas, da base de dados de Ferreira (2019), uma vez que as bases de dados possuíam o atributo Estatura do empregado (altura) como atributo comum (Item 12).

No Quadro 7, apresentam-se a descrição, os tipos e a origem dos atributos utilizados na rede SOM, MLP 1 e MLP 2.

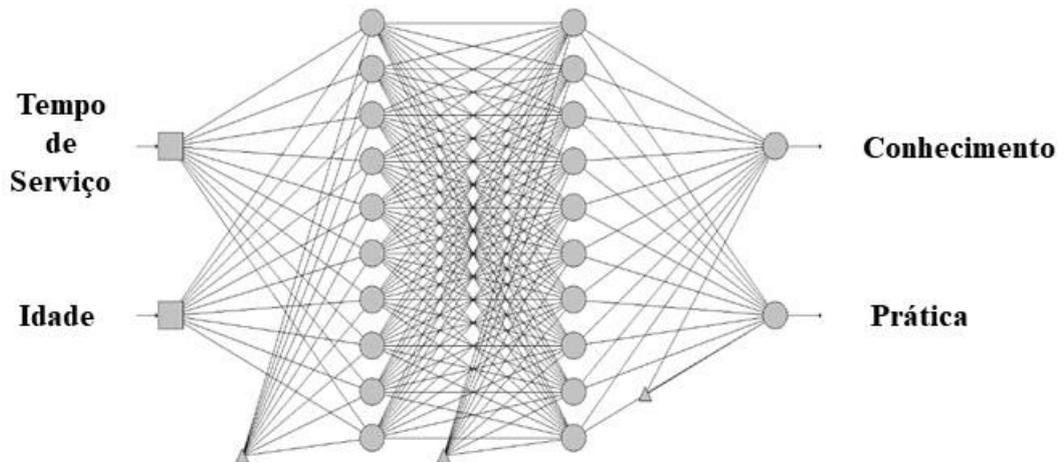
Quadro 7 – Descrição dos atributos utilizados na rede SOM, MLP 1 e MLP 2

Item	Atributos usados para enriquecimento da base de dados por meio da rede SOM	Atributos usados para enriquecimento da base de dados por meio da rede MLP 1	Atributos usados para enriquecimento da base de dados por meio da rede MLP 2	Descrição dos atributos dos experimentos computacionais (enriquecimento)	Tipo de atributo	Origem do atributo por base de dados
1				Corresponde ao código individual	Preditivo	Ferreira (2019)
2				Corresponde aos motivos do absentéismo	Preditivo	Ferreira (2019)
3				Corresponde ao tempo de permanência do indivíduo na base de dados em anos	Preditivo	Ferreira (2019)
4				Corresponde ao dia da semana	Preditivo	Ferreira (2019)
5				Corresponde ao tempo de afastamento em dias e horas, medido e transformado em horas	Preditivo	Ferreira (2019)
6				Corresponde à quantidade de licenças médicas registradas por mês	Preditivo	Ferreira (2019)
7		X		Corresponde ao tempo de serviço em anos completos	Preditivo	Ferreira (2019)
8		X		Corresponde à idade em anos completos	Preditivo	Ferreira (2019)
9				Corresponde à frequência das licenças médicas por indivíduo	Preditivo	Ferreira (2019)
10				Corresponde à classificação do indivíduo (presenteísta, normal e absenteísta)	Alvo (Previsão)	Ferreira (2019)
11	X			Corresponde ao peso em quilograma	Preditivo	Ferreira (2019)
12	X		X	Corresponde à estatura do empregado em centímetros (altura/entrada)	Preditivo	Defani (2007)
13	X		X	Corresponde ao comprimento do braço	Alvo MLP 2	Defani (2007)
14	X		X	Corresponde à altura dos ombros (do pé até o ombro)	Alvo MLP 2	Defani (2007)
15	X		X	Corresponde à altura do cotovelo (do pé até o cotovelo)	Alvo MLP 2	Defani (2007)
16	X		X	Corresponde à altura do púbis (do pé até a região pubiana)	Alvo MLP 2	Defani (2007)
17	X		X	Corresponde à altura do epigástrico (do pé até a altura da região epigástrica)	Alvo MLP 2	Defani (2007)
18	X			Corresponde ao IMC – Valores contínuos	Preditivo	Ferreira (2019)
19	X			Corresponde ao IMC – Valores categóricos	Preditivo	Ferreira (2019)
20	X			Corresponde ao tamanho do pé	Preditivo	Ferreira (2019)
21	X			Corresponde ao tamanho do tronco	Preditivo	Ferreira (2019)
22	X			Corresponde ao tamanho da cintura	Preditivo	Ferreira (2019)
23	X	X		Corresponde ao conhecimento sobre ergonomia	Alvo MLP 1	Martiniano, Ferreira e Sassi (2018)
24	X	X		Corresponde à prática de ergonomia no dia a dia	Alvo MLP 1	Martiniano, Ferreira e Sassi (2018)
25				Características antropométricas e ergonômicas (Rede SOM)	Autor	Autor

Fonte: Elaborado pelo autor (2020).

Na Figura 12, ilustram-se a topologia da MLP 1 e os atributos de entrada (Preditivos) e saída (Alvos) utilizados para o enriquecimento da base de dados. Ambos se referem aos Itens 23 e 24 respectivamente, apresentados no Quadro 5.

Figura 12 – Topologia da MLP 1

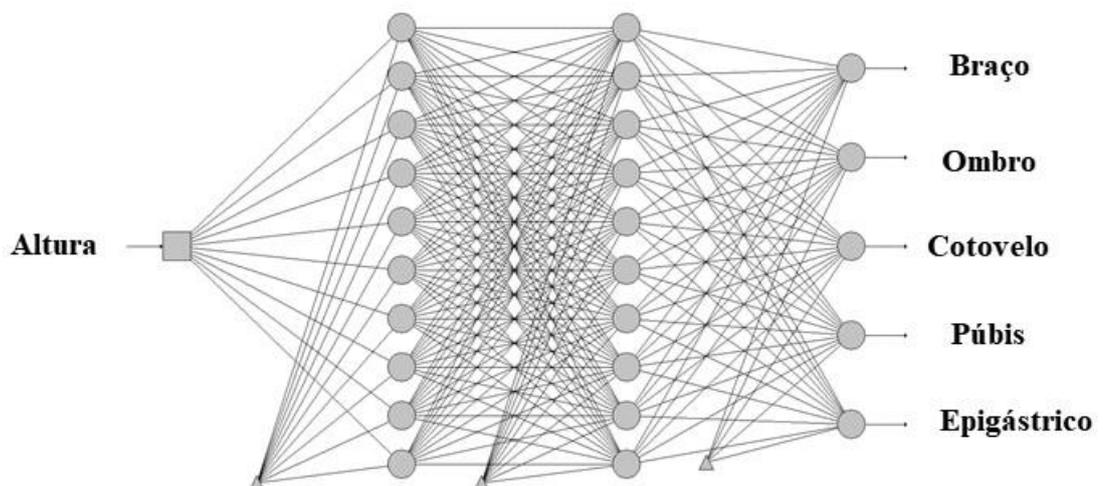


Fonte: Elaborado pelo autor (2020).

As entradas da MLP 1 foram os atributos Tempo de serviço e Idade em anos, sendo ambos comuns às bases de dados de Ferreira (2019) e Martiniano, Ferreira e Sassi (2018). Já as saídas da MLP 1 foram os atributos Conhecimento e Prática de ergonomia.

Na Figura 13, ilustram-se a topologia da MLP 2 e os atributos de entrada (Preditivo) e saída (Alvos) utilizadas para o enriquecimento da base de dados. Ambos se referem aos Itens 13, 14, 15, 16 e 17 respectivamente, apresentados do Quadro 5.

Figura 13 – Topologia da MLP 2



Fonte: Elaborado pelo autor (2020).

A entrada da MLP 2 foi a altura (estatura) dos empregados, sendo esse atributo comum às bases de dados de Ferreira (2019) e Defani (2007). As saídas da MLP 2 foram os atributos Comprimento do braço, Altura dos ombros (do pé até o ombro), Altura do cotovelo (do pé até o cotovelo), Altura do púbis (do pé até a região pubiana) e Altura do epigástrico (do pé até a altura da região epigástrica).

3.3.3 Bases de dados utilizadas para auxiliar na previsão do comportamento no trabalho

Após o enriquecimento, as três bases de dados foram consolidadas a fim de serem utilizadas nos experimentos computacionais para auxiliar na previsão do comportamento dos empregados no trabalho.

A base de dados de número 1 possui 10 atributos e 2.403 registros. A base de dados de número 2 possui 11 atributos e 2.403 registros. Já a base de dados de número 3 possui 25 atributos e 2.403 registros.

No Quadro 8, apresenta-se a descrição dos atributos utilizados nos experimentos computacionais para auxiliar na previsão do comportamento no trabalho. Vale ressaltar que o atributo que corresponde à classificação do indivíduo (presenteísta, normal e absenteísta) (Item 10) é o atributo-alvo das bases de dados 1, 2 e 3.

Quadro 8 – Descrição dos atributos para auxílio à previsão do comportamento no trabalho

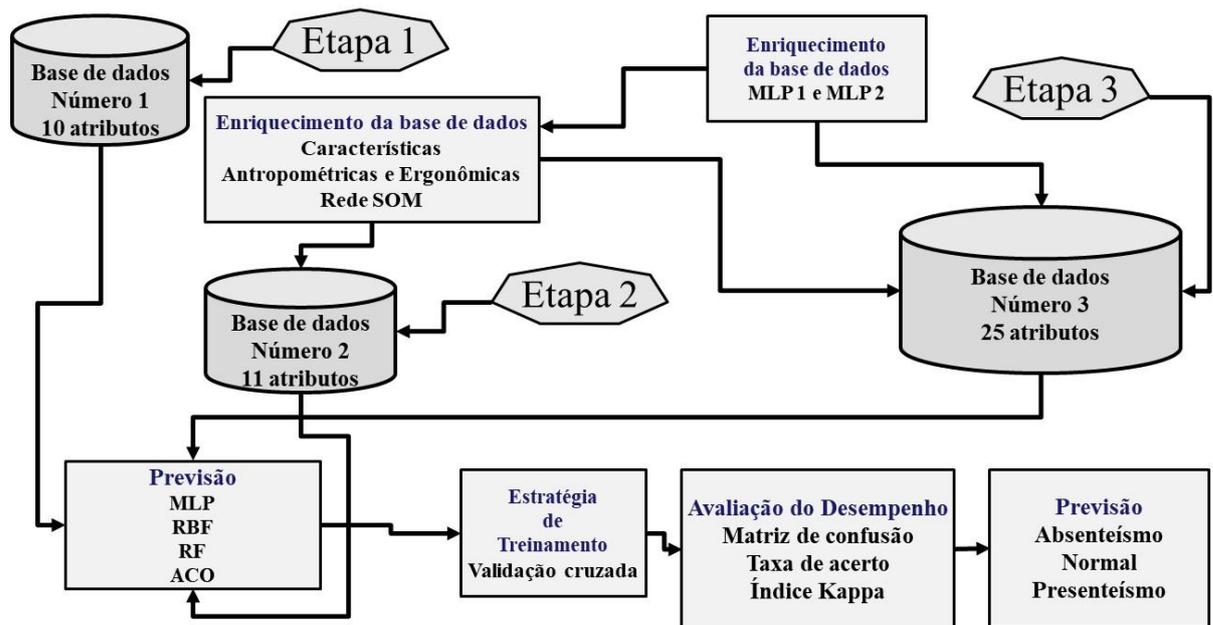
Item	Base de Dados Nº1	Base de Dados Nº2	Base de Dados Nº3	Descrição dos atributos dos experimentos computacionais (previsão)	Tipo de atributo	Origem do atributo por base de dados
1	X	X	X	Corresponde ao código individual	Preditivo	Ferreira (2019)
2	X	X	X	Corresponde aos motivos do absentéismo	Preditivo	Ferreira (2019)
3	X	X	X	Corresponde ao tempo de permanência do indivíduo na base de dados em anos	Preditivo	Ferreira (2019)
4	X	X	X	Corresponde ao dia da semana	Preditivo	Ferreira (2019)
5	X	X	X	Corresponde ao tempo de afastamento em dias e horas, medido e transformado em horas	Preditivo	Ferreira (2019)
6	X	X	X	Corresponde à quantidade de licenças médicas registradas por mês	Preditivo	Ferreira (2019)
7	X	X	X	Corresponde ao tempo de serviço em anos completos	Preditivo	Ferreira (2019)
8	X	X	X	Corresponde à idade em anos completos	Preditivo	Ferreira (2019)
9	X	X	X	Corresponde à frequência das licenças médicas por indivíduo	Preditivo	Ferreira (2019)
10	X	X	X	Corresponde à classificação do indivíduo (presenteísta, normal e absenteísta)	Alvo	Ferreira (2019)
11			X	Corresponde ao peso em quilograma	Preditivo	Ferreira (2019)
12			X	Corresponde à estatura do empregado em centímetros (altura/entrada)	Preditivo	Defani (2007)
13			X	Corresponde ao comprimento do braço	Preditivo	Defani (2007)
14			X	Corresponde à altura dos ombros (do pé até o ombro)	Preditivo	Defani (2007)
15			X	Corresponde à altura do cotovelo (do pé até o cotovelo)	Preditivo	Defani (2007)
16			X	Corresponde à altura do púbis (do pé até a região pubiana)	Preditivo	Defani (2007)
17			X	Corresponde à altura do epigástrico (do pé até a altura da região epigástrica)	Preditivo	Defani (2007)
18			X	Corresponde ao IMC – Valores contínuos	Preditivo	Ferreira (2019)
19			X	Corresponde ao IMC – Valores categóricos	Preditivo	Ferreira (2019)
20			X	Corresponde ao tamanho do pé	Preditivo	Ferreira (2019)
21			X	Corresponde ao tamanho do tronco	Preditivo	Ferreira (2019)
22			X	Corresponde ao tamanho da cintura	Preditivo	Ferreira (2019)
23			X	Corresponde ao conhecimento sobre ergonomia	Preditivo	Martiniano, Ferreira e Sassi (2018)
24			X	Corresponde à prática de ergonomia no dia a dia	Preditivo	Martiniano, Ferreira e Sassi (2018)
25		X	X	Características antropométricas e ergonômicas (Rede SOM)	Preditivo	Autor

Fonte: Elaborado pelo autor (2020).

3.4 CONDUÇÃO DOS EXPERIMENTOS COMPUTACIONAIS

Os experimentos computacionais para auxiliar na previsão do comportamento dos empregados no trabalho, que pode ser classificado como presenteísta, normal e absenteísta, foram conduzidos em três etapas, conforme exemplificado na Figura 14, com as bases de dados 1, 2 e 3, contendo os atributos 10, 11 e 25, respectivamente.

Figura 14 – Etapas dos experimentos computacionais



Fonte: Elaborado pelo autor (2020).

Etapa 1 – Foram realizados experimentos com as técnicas MLP, RBF, RF e ACO, utilizando a base de dados com 10 atributos (Base de dados Número 1).

Etapa 2 – Foram realizados experimentos com as técnicas MLP, RBF, RF e ACO, utilizando a base de dados enriquecida que contém 11 atributos (Base de dados Número 2).

Etapa 3 – Serão realizados experimentos com as técnicas MLP, RBF, RF e ACO, utilizando a base de dados enriquecida que contém 25 atributos (Base de dados Número 3). Para auxílio à previsão do comportamento no trabalho, foi utilizada a validação cruzada como estratégia de treinamento com três conjuntos de dados nas Etapas 1 e 2 e 3.

3.4.1 Parâmetros utilizados nas técnicas aplicadas

Os parâmetros utilizados nas técnicas aplicadas são apresentados no Quadro 9 a seguir.

Quadro 9 – Parâmetros utilizados nas técnicas

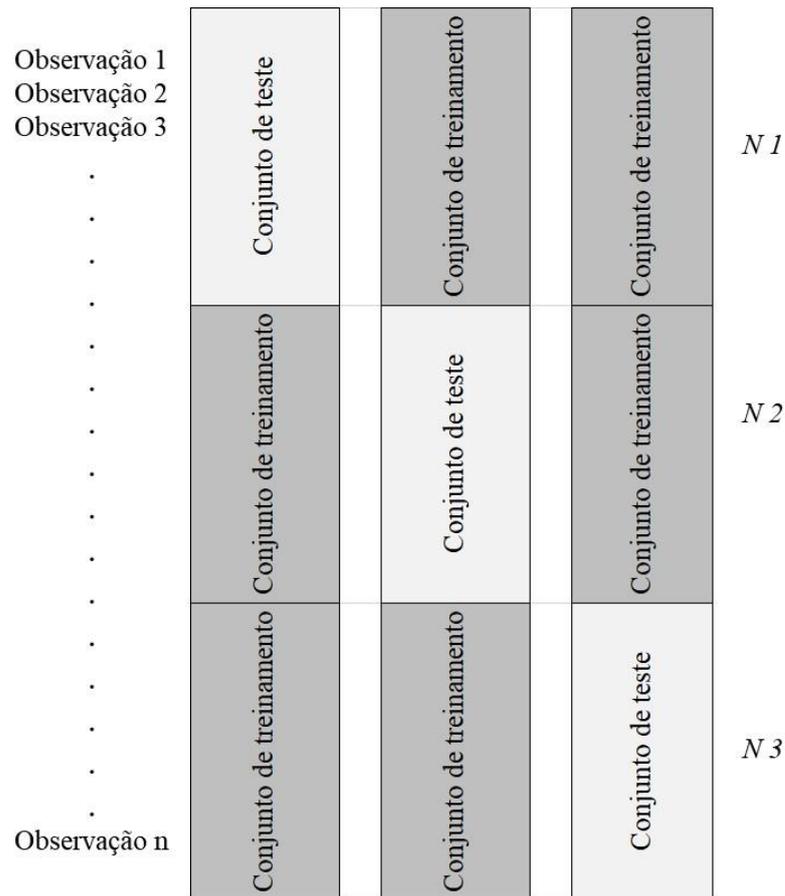
Técnica	Parâmetros utilizados
MLP	Número de neurônios na camada de entrada, número de camadas ocultas, número de neurônios em cada camada oculta, número máximo de iterações, como critério de parada, taxa de aprendizado, fator de momento e função de ativação.
RBF	Número de neurônios na camada de entrada, número de camadas ocultas, número de neurônios em cada camada oculta, número máximo de iterações, como critério de parada, taxa de aprendizado, fator de momento e função de ativação.
SOM	Número de neurônios, vizinhança topológica, função de vizinhança topológica, número máximo de iterações e taxa de aprendizado.
RF	Número de árvores da floresta, número mínimo de dados em cada nó terminal e número de variáveis utilizadas em cada árvore.
ACO	Exemplos mínimos cobertos por valor da regra, tamanho do teste de convergência, número de formigas na colônia, número máximo de iterações, como critério de parada.

Fonte: Elaborado pelo autor (2020).

3.4.2 Estratégia de treinamento: validação cruzada (*k-fold cross validation*)

A validação cruzada, ou *k-fold cross validation*, é uma estratégia de treinamento gerada por uma técnica de amostragem. Trata-se de um método que está baseado na divisão do conjunto de dados em N partições, mutuamente exclusivas, de tamanho igual. O modelo é gerado a partir de $N-1$ partes, e testado na parte que foi removida do conjunto de treinamento. Esse procedimento é repetido N vezes, até que todas as partes tenham sido usadas para teste do modelo (WITTEN et al., 2016). Na Figura 15, ilustra-se a validação cruzada, usando N partições.

Vale ressaltar que para a preparação dos dados não foi realizado pré-processamento, limpeza dos dados ou adequação do formato dos dados. Os dados selecionados para integrar a base de dados dos experimentos foram coletados de diferentes fontes, enriquecendo o conjunto de dados que foi utilizado nos experimentos computacionais.

Figura 15 – Exemplo de validação cruzada, usando N partições

Fonte: Adaptado de Xiong et al. (2020).

O valor de N está atrelado à quantidade total de amostras disponíveis, e o desempenho global do modelo será obtido em função da média entre os desempenhos individuais observados quando da aplicação das N partições (SILVA; SPATTI; FLAUZINO, 2016).

3.4.3 Medidas de avaliação do desempenho

a) Matriz de confusão

A matriz de confusão quantifica quantos exemplos do conjunto foram classificados corretamente e incorretamente. Além disso, permite visualizar não somente a taxa de acerto total de um dado classificador, mas seu comportamento, ao classificar os exemplos em cada classe.

Em um problema com duas classes, denominadas classe positiva e classe negativa, a matriz de confusão indica as quatro possibilidades de acertos e erros do classificador:

- Verdadeiros Positivos (VP): quando os exemplos de valor correto forem classificados como correto.

- Falsos Negativos (FN): quando os exemplos de valor correto forem classificados como incorreto.

- Verdadeiros Negativos (VN): quando os exemplos de valor incorreto forem classificados como incorreto.

- Falsos Positivos (FP): quando os exemplos de valor incorreto forem classificados como correto. Na Tabela 4, é exemplificada uma matriz de confusão com as quatro possibilidades.

Tabela 4 – Exemplo de uma matriz de confusão

	Classificação Correta	Classificação incorreta
Valor correto	VP	FN
Valor incorreto	FP	VN

Fonte: Adaptado de Witten et al. (2016).

b) Taxa de acerto

A taxa de acerto é uma medida de avaliação básica para modelos de classificação, ou seja, é o percentual de acertos no modelo de classificação. A taxa de acerto pode ser calculada a partir da matriz de confusão (WITTEN et al., 2016). Na Equação (7), demonstra-se o cálculo da taxa de acerto.

$$Taxa_de_acerto = \frac{VP + VN}{n} \quad (7)$$

Na qual, n é o número total de exemplos no conjunto.

c) Índice Kappa

O índice Kappa é usado para descrever a concordância entre dois ou mais jurados, ou entre dois métodos de classificação (SIEGEL; CASTELLAN, 1988; BANERJEE et. al., 1999; FLEISS; LEVIN; PAIK, 2003). O cálculo do índice Kappa leva em consideração, além da concordância percentual entre os jurados, a concordância que seria esperada ao acaso (WITTEN et al., 2016). Quando os jurados concordam em todas as observações, o valor do índice Kappa é igual a 1. Por outro lado, caso a concordância entre os jurados seja igual à concordância esperada ao acaso, o índice Kappa é igual a zero.

Para uma matriz de confusão, o índice Kappa (*IK*) pode ser calculado por meio da determinação da Probabilidade esperada (*Pe*), que é a possibilidade de ocorrer correlação ao acaso, e da Probabilidade observada (*Po*), que é a correlação que realmente ocorreu. Calculada pelas Equações (8), (9) e (10):

Cálculo da Probabilidade esperada (*Pe*):

$$Pe = \frac{(VP + FN) \times (VP + FP) + (VN + FP) \times (VN + FN)}{n^2} \quad (8)$$

Cálculo da Probabilidade observada (*Po*):

$$Po = \frac{(VP + VN)}{(VP + FN + FP + VN)} \quad (9)$$

Cálculo do Índice Kappa (*IK*):

$$IK = \frac{(Po - Pe)}{(1 - Pe)} \quad (10)$$

O índice Kappa é uma medida que assume valores entre -1 e 1, onde 1 significa concordância perfeita, e -1, uma discordância perfeita. Valores intermediários representam concordância em outros níveis. Na Tabela 5, mostra-se a interpretação dos valores do índice Kappa proposta por Landis e Koch (1977).

Tabela 5 – Interpretação dos valores do índice Kappa

Intervalo	Nível de concordância dos valores do índice Kappa
<0	Ausência de concordância
0-0,19	Concordância pobre
0,20-0,39	Concordância leve
0,40-0,59	Concordância moderada
0,60-0,79	Concordância substantiva
0,80-1,00	Concordância quase perfeita

Fonte: Adaptado de Landis e Koch (1977).

d) 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 (w) do neurônio BMU (*Best Match Unit*). O BMU é calculado por meio da Equação (11).

$$\arg \min_j \left(\|w_j - x_i\| \right) \quad (11)$$

A medida EQ corresponde à resolução do mapa. Valores menores indicam um ajuste melhor. Kohonen (2001) sugeriu o EQ como uma medida básica de qualidade para avaliar a auto-organização do mapa. Para calcular o EQ, utiliza-se a Equação (12) (BREARD, 2017).

$$EQ = \frac{1}{n} \sum_{i=1}^n \|x_i - w_i\| \quad (12)$$

Estima-se que quanto menor o EQ, mais ajustado o neurônio vencedor estará aos vetores de entrada (BREARD, 2017).

- O Erro Topográfico (ET) quantifica a capacidade do mapa em representar a topologia dos dados de entrada da rede SOM. Para calcular o ET, utiliza-se a Equação (13).

$$ET = \frac{1}{N} \sum_{k=1}^N u(x_k) \quad (13)$$

No ET, 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 vizinhos, aumenta-se o erro em uma unidade, tirando-se depois a média pelo número total de vetores, em que $u(x_k)$ é igual a 1 se o primeiro e o segundo BMU não forem vizinhos, caso contrário, será igual a 0 (BREARD, 2017).

Um dos principais objetivos da rede SOM é preservar a topologia do espaço de entrada n dimensional no espaço de saída bidimensional. O ET é uma medida de quão bem a estrutura do espaço de entrada é modelada pelo mapa de saída, ou seja, avalia as discontinuidades locais no mapeamento (KIVILUOTO, 1996; BREARD, 2017).

Segundo Kohonen (2001), o EQ corresponde à resolução do mapa. Estima-se que quanto menor o EQ, mais bem ajustado o neurônio vencedor estará aos vetores de entrada. Entretanto, quando os valores do EQ e ET são muito próximos de zero, suspeita-se de *overfitting* (BREARD, 2017).

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Neste capítulo, são apresentados e discutidos os resultados dos três experimentos computacionais, realizados para o enriquecimento das bases de dados, sendo que um foi desenvolvido com a rede SOM, e os outros dois, com a MLP. Para auxílio à previsão do comportamento no trabalho, foram realizados também quatro experimentos com as seguintes técnicas: MLP, RBF, RF e ACO.

4.1 EXPERIMENTOS COMPUTACIONAIS REALIZADOS PARA ENRIQUECIMENTO DAS BASES DE DADOS

a) Aplicação da rede SOM

As bases de dados 2 e 3 foram enriquecidas com o resultado da aplicação da rede SOM, utilizando os atributos antropométricos e ergonômicos, oriundos de Ferreira (2019), Defani (2007) e Martiniano, Ferreira e Sassi (2018), e serviram para as bases de dados dos experimentos computacionais das Etapas 2 e 3. O atributo gerado pela rede SOM, que enriqueceu as bases de dados, foi chamado de Características antropométricas e ergonômicas.

A rede SOM foi utilizada para agrupar as características antropométricas e ergonômicas em um único atributo com três classes (agrupamentos). No Quadro 10, apresentam-se os 14 atributos aplicados na rede SOM.

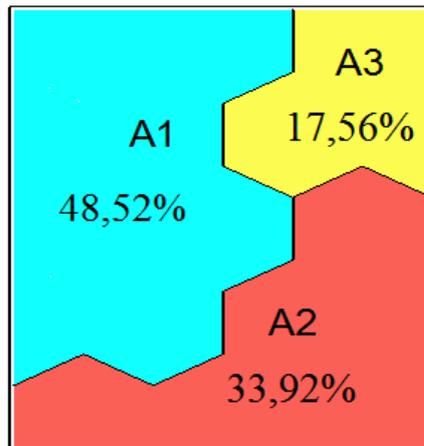
Quadro 10 – Atributos aplicados na rede SOM (14 atributos)

Item	Atributos usados para enriquecimento das bases de dados por meio da rede SOM
11	Corresponde ao peso em quilograma
12	Corresponde à estatura do empregado em centímetros (altura/entrada)
13	Corresponde ao comprimento do braço
14	Corresponde à altura dos ombros (do pé até o ombro)
15	Corresponde à altura do cotovelo (do pé até o cotovelo)
16	Corresponde à altura do púbis (do pé até a região pubiana)
17	Corresponde à altura do epigástrico (do pé até a altura da região epigástrica)
18	Corresponde ao IMC – Valores contínuos
19	Corresponde ao IMC – Valores categóricos
20	Corresponde ao tamanho do pé
21	Corresponde ao tamanho do tronco
22	Corresponde ao tamanho da cintura
23	Corresponde ao conhecimento sobre ergonomia
24	Corresponde à prática de ergonomia no dia a dia

Fonte: Elaborado pelo autor (2020).

Na Figura 16, ilustram-se os agrupamentos gerados pela rede SOM.

Figura 16 – Agrupamentos gerados pela rede SOM



Fonte: Elaborado pelo autor (2020).

O agrupamento A1 contém 19 empregados, o A2, 13, e o A3, 7, totalizando 39 empregados. O agrupamento A1 representa 48,52% dos registros, o A2, 33,92%, e o A3, 17,56%.

Os parâmetros utilizados na topologia da rede SOM foram: número de neurônios igual a 25 (5x5), vizinhança topológica hexagonal e função de vizinhança topológica gaussiana. Os parâmetros utilizados na fase de treinamento foram: número de épocas igual a 3000 e taxa de aprendizado igual a 0,5, na fase inicial, e 0,05, na fase de convergência. O EQ foi de 0,291869, e o ET foi de 0,095963, ambos considerados bons resultados para a formação de agrupamentos.

b) Aplicação da MLP 1

As bases de dados 2 e 3 foram enriquecidas com os resultados da aplicação da MLP 1, utilizando os atributos sobre ergonomia de Martiniano, Ferreira e Sassi (2018) e Ferreira (2019), e serviram para enriquecer a base de dados dos experimentos das Etapas 2 e 3.

Os atributos gerados pela MLP 1 foram: conhecimento sobre ergonomia e prática de ergonomia no dia a dia. Na MLP 1, foram aplicados como entrada os atributos Tempo de serviço e Idade. Na Tabela 6, são apresentados os parâmetros utilizados na MLP 1.

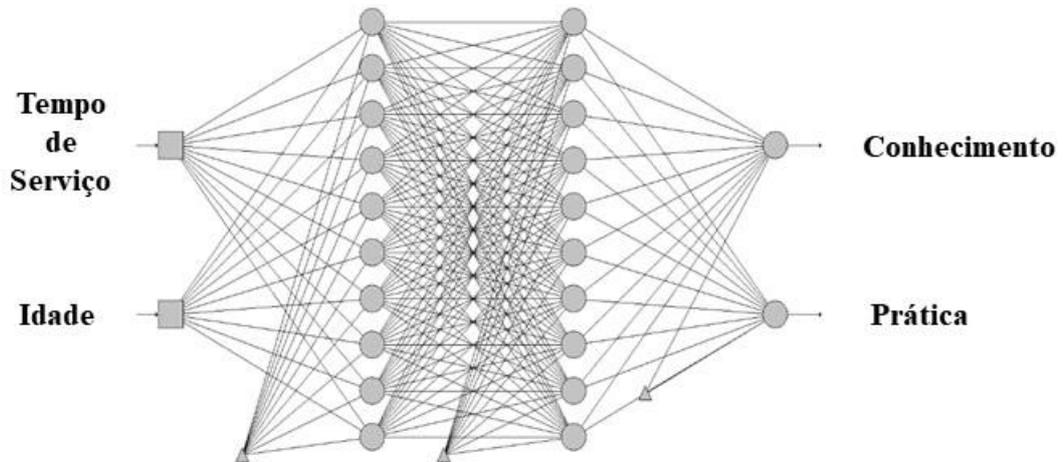
Tabela 6 – Parâmetros utilizados na MLP 1 (dados de ergonomia)

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios nas camadas ocultas	Taxa inicial de aprendizado	Decaimento a cada 100 épocas	Fator de momento inicial	Decaimento a cada 200 épocas
2	2	10	0,7	1%	0,7	1%

Fonte: Elaborado pelo autor (2020).

O número máximo de iterações foi igual a 1500. A função de ativação dos neurônios das camadas ocultas e das saídas da MLP 1 foi a sigmoide. Na Figura 17, ilustra-se a topologia da MLP 1 utilizada no experimento.

Figura 17 – Topologia da MLP 1 utilizada no experimento



Fonte: Elaborado pelo autor (2020).

c) Aplicação da MLP 2

As bases de dados 2 e 3 foram enriquecidas com os resultados da aplicação da MLP 2, utilizando os atributos antropométricos oriundos de Defani (2007) e Ferreira (2019), e serviram para a base de dados dos experimentos das Etapas 2 e 3. Os atributos gerados pela MLP 2 foram:

- Comprimento do braço;
- Altura dos ombros (do pé até o ombro);
- Estatura do empregado (altura);
- Altura do cotovelo (do pé até o cotovelo);
- Altura do púbis (do pé até a região pubiana);
- Altura do epigástrico (do pé até a altura da região epigástrica).

Na MLP 2, foi aplicado como entrada o atributo Altura. Na Tabela 7, são apresentados os parâmetros utilizados na MLP 2.

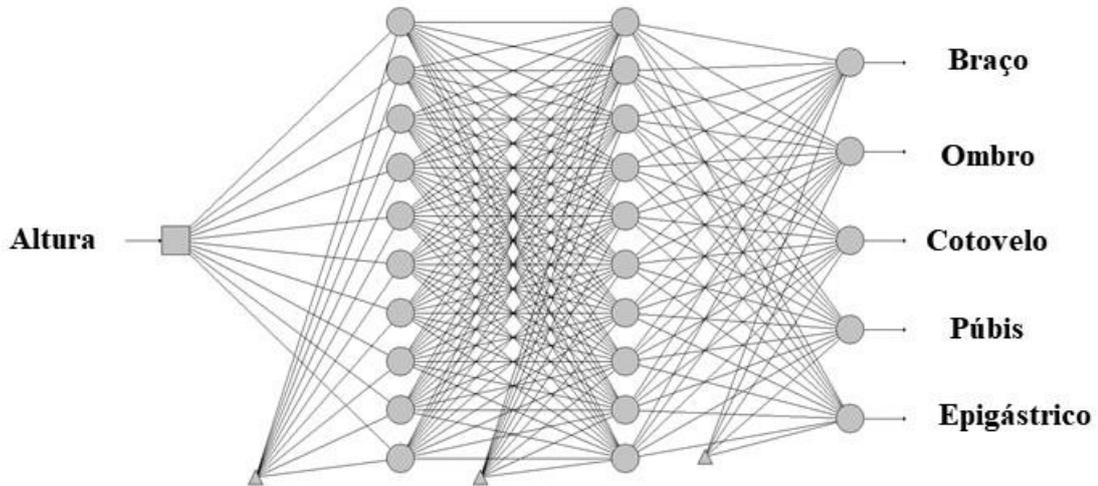
Tabela 7 – Parâmetros utilizados na MLP 2 (dados antropométricos)

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios nas camadas ocultas	Taxa inicial de aprendizado	Decaimento a cada 100 épocas	Fator de momento inicial	Decaimento a cada 200 épocas
1	2	10	0,7	1%	0,7	1%

Fonte: Elaborado pelo autor (2020).

O número máximo de iterações foi igual a 1500. A função de ativação dos neurônios das camadas ocultas e das saídas da MLP 2 foi a sigmoide. Na Figura 18, ilustra-se a topologia da MLP 2 utilizada no experimento.

Figura 18 – Topologia da MLP 2 utilizada no experimento.



Fonte: Elaborado pelo autor (2020).

4.2 REALIZAÇÃO DOS EXPERIMENTOS COMPUTACIONAIS

Como descrito no capítulo 3, a realização dos experimentos com as técnicas MLP, RBF, RF e ACO foi dividida em três etapas distintas:

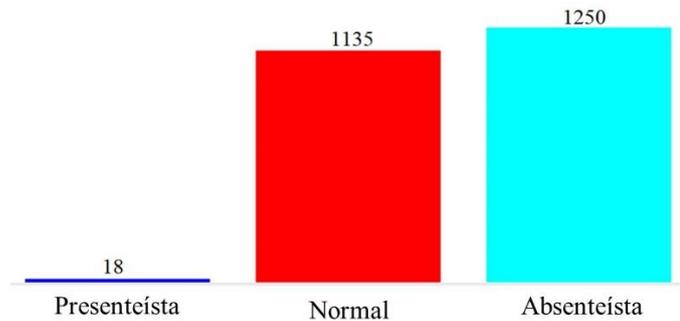
Etapa 1 – Foram realizados experimentos para auxílio à previsão dos comportamentos presenteísta, normal e absenteísta, com MLP, RBF, RF e ACO, utilizando a base de dados com 10 atributos (Base de dados Número 1).

Etapa 2 – Foram realizados experimentos para auxílio à previsão dos comportamentos presenteísta, normal e absenteísta, com MLP, RBF, RF e ACO, utilizando a base de dados enriquecida que contém 11 atributos (Base de dados Número 2).

Etapa 3 – Foram realizados experimentos para auxílio à previsão dos comportamentos presenteísta, normal e absenteísta, com MLP, RBF, RF e ACO, utilizando a base de dados enriquecida que contém 25 atributos (Base de dados Número 3). Para auxílio à previsão do comportamento no trabalho, foi utilizada a validação cruzada como estratégia de treinamento com três conjuntos de dados nas Etapas 1, 2 e 3.

Na Figura 19, apresenta-se a quantidade de registros nas bases de dados com comportamentos presenteísta, normal e absenteísta, que são as classes utilizadas pelas técnicas selecionadas.

Figura 19 – Quantidade de registros com comportamentos presenteísta, normal e absenteísta



Fonte: Elaborado pelo autor (2020).

Os experimentos computacionais referentes ao auxílio à previsão do comportamento dos empregados são descritos a seguir.

4.2.1 Experimentos computacionais – Etapa 1 – Base de dados Número 1 com 10 atributos

- Previsão com a **MLP** aplicada na Etapa 1

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios das camadas ocultas e da saída da MLP foi a sigmoide. A saída da MLP foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho. Na Tabela 8, apresentam-se os parâmetros utilizados na MLP aplicada na Etapa 1.

Tabela 8 – Parâmetros utilizados na MLP aplicada na Etapa 1

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios nas camadas ocultas	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
9	2	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão com a MLP, aplicada na Etapa 1, apresentou uma taxa de acerto de 99,91% de registros classificados corretamente. O índice Kappa foi de 0,9984, o que indica uma concordância quase perfeita do modelo. A MLP classificou incorreto apenas 2 registros. Na Figura 20, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 20 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP na Etapa 1

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	16	2	0
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a **RBF** aplicada na Etapa 1

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios da camada oculta foi a gaussiana, e da saída da RBF foi a linear. A saída da RBF foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho. Na Tabela 9, são apresentados os parâmetros utilizados na RBF aplicada na Etapa 1.

Tabela 9 – Parâmetros utilizados na RBF aplicada na Etapa 1

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios na camada oculta	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
9	1	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão com a RBF, aplicada na Etapa 1, apresentou uma taxa de acerto de 97,08% de registros classificados corretamente. O índice Kappa foi de 0,9425, o que indica uma concordância quase perfeita do modelo. A RBF classificou incorreto 70 registros. Na Figura 21, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 21 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RBF na Etapa 1

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	6	8	4
Normal	6	1104	25
Absenteísta	7	20	1223

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação da **RF** na Etapa 1

Os parâmetros usados na RF foram: o número de árvores da floresta igual a 100, o número mínimo de dados em cada nó terminal igual a 5 e o número de variáveis utilizadas em cada árvore igual a 3 (BREIMAN, 2001). A saída da RF foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A previsão com a RF, aplicada na Etapa 1, apresentou uma taxa de acerto de 99,91% de registros classificados corretamente. O índice Kappa foi de 0,9984, o que indica uma concordância quase perfeita do modelo. A RF classificou incorreto apenas 2 registros. Na Figura 22, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 22 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF na Etapa 1

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	16	1	1
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação do **ACO** na Etapa 1

Os parâmetros utilizados no ACO foram: exemplos mínimos cobertos por valor da regra igual a 10, tamanho do teste de convergência igual a 10, número de formigas na colônia igual

a 60, e número máximo de iterações, como critério de parada, igual a 1500. A saída do ACO foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A previsão com o ACO, aplicada na Etapa 1, apresentou uma taxa de acerto de 80,65% de registros classificados corretamente. O índice Kappa foi de 0,6127, o que indica uma concordância substantiva do modelo. O ACO classificou incorreto 465 registros. Na Figura 23, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 23 – Matriz de confusão com as classificações corretas e incorretas da aplicação do ACO na Etapa 1

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	11	7	0
Normal	0	775	360
Absenteísta	0	98	1152

Fonte: Elaborado pelo autor (2020).

Os melhores desempenhos na Etapa 1 foram obtidos com a aplicação das técnicas MLP e RF, pois ambas apresentaram uma taxa de acerto de 99,91% de registros classificados corretamente. Posteriormente, quem apresentou melhor desempenho foi a RBF, com uma taxa de acerto de 97,08%, seguida do ACO, com 80,65% de registros classificados corretamente.

No Quadro 11, apresenta-se uma síntese dos experimentos computacionais, considerando também o tempo de processamento da Etapa 1.

Quadro 11 – Síntese dos experimentos da Etapa 1

	Etapa 1 - 10 atributos (Base de dados Número 1)		
Classificadores	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)
MLP	99,91	0,9984	186,93
RBF	97,08	0,9425	0,45
RF	99,91	0,9984	0,48
ACO	80,65	0,6127	20

Fonte: Elaborado pelo autor (2020).

4.2.2 Experimentos computacionais – Etapa 2 – Base de dados Número 2 com 11 atributos

- Previsão com a aplicação da **MLP** na Etapa 2

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios das camadas ocultas e da saída da MLP foi a sigmoide. A saída da MLP foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho. Na Tabela 10, apresentam-se os parâmetros utilizados na MLP aplicada na Etapa 2.

Tabela 10 – Parâmetros utilizados na MLP aplicada na Etapa 2

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios nas camadas ocultas	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
10	2	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão com a MLP, aplicada na Etapa 2, apresentou uma taxa de acerto de 99,91% de registros classificados corretamente. O índice Kappa foi de 0,9984, o que indica uma concordância quase perfeita do modelo. A MLP classificou incorreto apenas 2 registros. Na Figura 24, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 24 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP na Etapa 2

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	16	2	0
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação da **RBF** na Etapa 2

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios da camada oculta foi a gaussiana, e da saída da RBF foi a linear. A saída da RBF foi a previsão do comportamento presenteísta, normal ou absenteísta. Na Tabela 11, são apresentados os parâmetros utilizados na RBF aplicada na Etapa 2.

Tabela 11 – Parâmetros utilizados na RBF aplicada na Etapa 2

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios na camada oculta	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
10	1	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão com a RBF, aplicada na Etapa 2, apresentou uma taxa de acerto de 97,25% de registros classificados corretamente. O índice Kappa foi de 0,9456, o que indica uma concordância quase perfeita do modelo. A RBF classificou incorreto 72 registros. Na Figura 25, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 25 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RBF na Etapa 2

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	6	9	3
Normal	5	1107	23
Absenteísta	0	26	1224

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação da RF na Etapa 2

Os parâmetros usados na RF foram: o número de árvores da floresta igual a 100, o número mínimo de dados em cada nó terminal igual a 5 e o número de variáveis utilizadas em cada árvore igual a 4 (BREIMAN, 2001). A saída da RF foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A classificação, com a aplicação da RF na Etapa 2, apresentou uma taxa de acerto de 99,91% de registros classificados corretamente. O índice Kappa foi de 0,9984, o que indica uma concordância quase perfeita do modelo. A RF classificou incorreto apenas 2 registros. Na Figura 26, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 26 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF na Etapa 2

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	16	2	0
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação do **ACO** na Etapa 2

Os parâmetros utilizados no ACO foram: exemplos mínimos cobertos por valor da regra igual a 10, tamanho do teste de convergência igual a 10, número de formigas na colônia igual a 60, número máximo de iterações, como critério de parada, igual a 1500. A saída do ACO foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A previsão com o ACO, aplicada na Etapa 2, apresentou uma taxa de acerto de 84,44% de registros classificados corretamente. O índice Kappa foi de 0,6885, o que indica uma concordância substantiva do modelo. O ACO classificou incorreto 374 registros. Na Figura 27, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 27 – Matriz de confusão com as classificações corretas e incorretas da aplicação do ACO na Etapa 2

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	12	6	0
Normal	0	815	320
Absenteísta	0	48	1202

Fonte: Elaborado pelo autor (2020).

Os melhores desempenhos na Etapa 2 foram obtidos com as técnicas MLP e RF, pois ambas apresentaram uma taxa de acerto de 99,91% de registros classificados corretamente. Posteriormente, quem apresentou melhor desempenho foi a RBF, com uma taxa de acerto de 97,25%, seguida do ACO, com uma taxa de acerto 84,44% de registros classificados

corretamente. Foi possível perceber uma sensível diminuição no tempo de processamento, com o acréscimo do atributo Características antropométricas e ergonômicas (Item 25), na Etapa 2, com o uso das técnicas MLP, RBF, RF e ACO, o que indica que, mesmo com o aumento dos dados para processamento, o atributo acrescido por meio do enriquecimento melhorou o modelo de previsão. Observou-se também que o tempo de processamento da MLP foi bastante elevado em comparação com a RBF, a RF e o ACO. No Quadro 12, apresenta-se uma síntese dos experimentos computacionais, considerando também o tempo de processamento da Etapa 2.

Quadro 12 – Síntese dos experimentos da Etapa 2

Classificadores	Etapa 2 - 11 atributos. (Base de dados Número 2)		
	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)
MLP	99,91	0,9984	179,82
RBF	97,25	0,9456	0,33
RF	99,91	0,9984	0,45
ACO	84,44	0,6885	18

Fonte: Elaborado pelo autor (2020).

No Quadro 13, faz-se uma comparação dos experimentos computacionais, considerando também o tempo de processamento das Etapas 1 e 2.

Quadro 13 – Comparação dos experimentos computacionais das Etapas 1 e 2

Classificadores	Etapa 1 - 10 atributos (Base de dados Número 1)			Etapa 2 - 11 atributos (Base de dados Número 2)		
	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)
MLP	99,91	0,9984	186,93	99,91	0,9984	179,82
RBF	97,08	0,9425	0,45	97,25	0,9456	0,33
RF	99,91	0,9984	0,48	99,91	0,9984	0,45
ACO	80,65	0,6127	20	84,44	0,6885	18

Fonte: Elaborado pelo autor (2020).

Os experimentos computacionais das Etapas 1 e 2, para a previsão do comportamento dos empregados no trabalho, mostraram que o enriquecimento da base de dados, com o atributo

Características ergonômicas e antropométricas, diminuiu o tempo de processamento das quatro técnicas aplicadas.

Além disso, percebeu-se que a RBF e o ACO apresentaram uma melhora no desempenho em relação à Etapa 1, além da diminuição do tempo de processamento. Observou-se ainda que, na Etapa 2, o ACO apresentou uma melhora mais significativa no desempenho.

Dessa forma, ficou evidente que o enriquecimento na base de dados, com o atributo Características ergonômicas e antropométricas, permitiu um melhor ajuste do modelo proposto na Etapa 2. É importante ressaltar que a MLP e a RF apresentaram uma taxa de acerto de 99,91%, respectivamente. Já a RBF apresentou uma taxa de acerto de 97,08%, e o ACO, de 80,65%, o que evidencia que as técnicas escolhidas foram adequadas para a resolução do problema de mineração de dados, uma vez que apresentaram resultados satisfatórios nos experimentos computacionais. Embora o ACO tenha apresentado um desempenho inferior às demais técnicas, sua taxa de acerto foi acima de 80%, e o índice Kappa foi de 0,6127, podendo ser interpretado como nível de concordância substantiva.

4.2.3 Experimentos computacionais – Etapa 3 – Base de dados Número 3 com 25 atributos

- Previsão com a aplicação da **MLP** na Etapa 3

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios das camadas ocultas e da saída da MLP foi a sigmoide. A saída da MLP foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho. Na Tabela 12, apresentam-se os parâmetros utilizados na MLP na Etapa 3.

Tabela 12 – Parâmetros utilizados na MLP aplicada na Etapa 3

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios nas camadas ocultas	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
10	2	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão, com a aplicação da MLP na Etapa 3, apresentou uma taxa de acerto de 99,96% de registros classificados corretamente. O índice Kappa foi de 0,9992, o que indica uma concordância quase perfeita do modelo. A MLP classificou incorreto apenas 1 registro. Na Figura 28, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 28 – Matriz de confusão com as classificações corretas e incorretas da aplicação da MLP na Etapa 3

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	17	1	0
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação da **RBF** na Etapa 3

O número máximo de iterações foi igual a 2500. A função de ativação dos neurônios da camada oculta foi a gaussiana, e da saída da RBF foi a linear. A saída da RBF foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho. Na Tabela 13, apresentam-se os parâmetros utilizados na RBF aplicada na Etapa 3.

Tabela 13 – Parâmetros utilizados na RBF aplicada na Etapa 3

Número de neurônios de entrada	Número de camadas ocultas	Número de neurônios na camada oculta	Taxa inicial de aprendizado	Decaimento a cada 200 épocas	Fator de momento inicial	Decaimento a cada 300 épocas
10	1	10	0,3	1%	0,2	1%

Fonte: Elaborado pelo autor (2020).

A previsão, mediante a aplicação da RBF na Etapa 3, apresentou uma taxa de acerto de 96,25% de registros classificados corretamente. O índice Kappa foi de 0,9259, o que indica uma concordância quase perfeita do modelo. A RBF classificou incorreto 90 registros. Na Figura 29, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 29 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RBF na Etapa 3

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	7	8	3
Normal	3	1097	35
Absenteísta	2	39	1209

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação da **RF** na Etapa 3

Os parâmetros usados na RF foram: o número de árvores da floresta igual a 100, o número mínimo de dados em cada nó terminal igual a 5 e o número de variáveis utilizadas em cada árvore igual a 5 (BREIMAN, 2001). A saída da RF foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A classificação, com a aplicação da RF na Etapa 3, apresentou uma taxa de acerto de 99,91% de registros classificados corretamente. O índice Kappa foi de 0,9984, o que indica uma concordância quase perfeita do modelo. A RF classificou incorreto apenas 2 registros. Na Figura 30, mostra a matriz de confusão com as classificações corretas e incorretas.

Figura 30 – Matriz de confusão com as classificações corretas e incorretas da aplicação da RF na Etapa 3

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	16	2	0
Normal	0	1135	0
Absenteísta	0	0	1250

Fonte: Elaborado pelo autor (2020).

- Previsão com a aplicação do **ACO** na Etapa 3

Os parâmetros utilizados no ACO foram: exemplos mínimos cobertos por valor da regra igual a 10, tamanho do teste de convergência igual a 10, número de formigas na colônia igual

a 60, e número máximo de iterações, como critério de parada, igual a 1500. A saída do ACO foi a previsão do comportamento presenteísta, normal ou absenteísta, no trabalho.

A previsão com o ACO, aplicada na Etapa 3, apresentou uma taxa de acerto de 91,80% de registros classificados corretamente. O índice Kappa foi de 0,8368, o que indica uma concordância quase perfeita do modelo. O ACO classificou incorreto 197 registros. Na Figura 31, mostra-se a matriz de confusão com as classificações corretas e incorretas.

Figura 31 – Matriz de confusão com as classificações corretas e incorretas da aplicação do ACO na Etapa 3

Classe	Presenteísta	Normal	Absenteísta
Presenteísta	15	3	0
Normal	0	955	180
Absenteísta	0	14	1236

Fonte: Elaborado pelo autor (2020).

No Quadro 14, apresenta-se uma síntese dos experimentos computacionais, considerando também o tempo de processamento da Etapa 3.

Quadro 14 – Síntese dos experimentos computacionais da Etapa 3

Classificadores	Etapa 3 - 25 atributos. (Base de dados Número 3)		
	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)
MLP	99,96	0,9992	269,61
RBF	96,25	0,9259	0,39
RF	99,91	0,9984	38,3
ACO	91,80	0,8368	36

Fonte: Elaborado pelo autor (2020).

No Quadro 15, faz-se uma comparação dos experimentos computacionais, considerando também o tempo de processamento das Etapas 1, 2 e 3.

Quadro 15 – Comparação dos experimentos computacionais das Etapas 1, 2 e 3

Classificadores	Etapa 1 - 10 atributos (Base de dados Número 1)			Etapa 2 - 11 atributos (Base de dados Número 2)			Etapa 3 - 25 atributos (Base de dados Número 3)		
	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)	Taxa de acerto (%)	Índice Kappa	Tempo de Processamento (s)
MLP	99,91	0,9984	186,93	99,91	0,9984	179,82	99,96	0,9992	269,61
RBF	97,08	0,9425	0,45	97,25	0,9456	0,33	96,25	0,9259	0,39
RF	99,91	0,9984	0,48	99,91	0,9984	0,45	99,91	0,9984	0,38
ACO	80,65	0,6127	20	84,44	0,6885	18	91,80	0,8368	36

Fonte: Elaborado pelo autor (2020).

Os experimentos computacionais das Etapas 1, 2 e 3, para a previsão do comportamento dos empregados no trabalho, mostraram que o enriquecimento da base de dados, com o atributo Características ergonômicas e antropométricas (Item 25), e os atributos sobre ergonomia e antropometria (Itens 11 a 24), melhorou seu desempenho com as técnicas MLP, RF e ACO. A RBF, por outro lado, foi uma exceção, pois apresentou menor desempenho na Etapa 3 em comparação com a Etapa 2. O tempo de processamento das quatro técnicas aplicadas foi maior do que na Etapa 2, com exceção da RF, que diminuiu o tempo de processamento em comparação com as Etapas 1 e 2. É importante ressaltar que, na Etapa 3, a MLP obteve o melhor desempenho com uma taxa de acerto de 99,96% e que a RF apresentou uma taxa de acerto de 99,91% com o tempo de processamento de 0,38 segundos. Já a RBF e o ACO apresentaram uma taxa de acerto acima de 90%.

A seguir, na seção 4.3, é apresentada uma análise detalhada dos resultados dos experimentos computacionais.

4.3 ANÁLISE DOS RESULTADOS DOS EXPERIMENTOS COMPUTACIONAIS

Nesta seção, descreve-se a análise dos resultados computacionais em cada etapa, bem como o desempenho de cada técnica (MLP, RBF, RF e ACO), considerando o tempo de processamento, as classes presenteísta, normal e absenteísta, e as matrizes de confusão.

4.3.1 Análise dos resultados computacionais da Etapa 1

No Quadro 16, faz-se uma comparação dos experimentos computacionais, levando em consideração o tempo de processamento, as classes de comportamento e as matrizes de confusão da Etapa 1.

Quadro 16 – Comparação dos experimentos computacionais da Etapa 1

	Classe	Presenteísta	Normal	Absenteísta	Taxa de acerto (%)	Índice Kappa	Tempo de processamento (s)
MLP	Presenteísta	16	2	0	99,91	0,9984	186,93
	Normal	0	1135	0			
	Absenteísta	0	0	1250			
RBF	Presenteísta	6	8	4	97,08	0,9425	0,45
	Normal	6	1104	25			
	Absenteísta	7	20	1223			
RF	Presenteísta	16	1	1	99,91	0,9984	0,48
	Normal	0	1135	0			
	Absenteísta	0	0	1250			
ACO	Presenteísta	11	7	0	80,65	0,6127	20
	Normal	0	775	360			
	Absenteísta	0	98	1152			

Fonte: Elaborado pelo autor (2020).

a) MLP

Conforme demonstrado no Quadro 16, observa-se que, na Etapa 1, o experimento computacional, com a aplicação da técnica MLP, apresentou uma taxa de acerto de 99,91%, um índice Kappa de 0,9984, e um tempo de processamento de 186,93 segundos. Na matriz de confusão, observa-se também que somente 2 registros foram classificados incorretamente. São resultados considerados muito bons por se aproximarem dos 100% de acerto, porém o tempo de processamento se mostrou elevado se comparado às demais técnicas.

b) RBF

Também conforme demonstrado no Quadro 16, observa-se que, na Etapa 1, o experimento computacional, com a aplicação da técnica RBF, apresentou uma taxa de acerto de 97,08%, um índice Kappa de 0,9425, e um tempo de processamento de 0,45 segundos. A RBF obteve um bom desempenho ao se aproximar de 100% na taxa de acerto e no índice. No

entanto, na matriz de confusão, 70 registros foram classificados de modo incorreto, e o tempo de processamento foi menor do que o da MLP, da RF e do ACO.

c) RF

Em relação à RF, observa-se que, na Etapa 1, o experimento computacional apresentou uma taxa de acerto de 99,91%, um índice Kappa de 0,9984, e um tempo de processamento de 0,48 segundos. Dessa forma, constata-se que a RF apresentou um desempenho muito bom na taxa de acerto e no índice Kappa, igualando-se ao desempenho da MLP. Na matriz de confusão, observa-se que foram classificados somente 2 registros incorretos, e que o tempo de processamento foi menor do que o da MLP e maior do que o da RBF.

d) ACO

Sobre o experimento computacional realizado com o ACO na etapa 1, observa-se que essa técnica apresentou uma taxa de acerto de 80,65%, um índice Kappa de 0,6127, e um tempo de processamento de 20 segundos. Sendo assim, fica evidente que o ACO apresentou um desempenho bem inferior ao das demais técnicas, tanto na taxa de acerto quanto no índice Kappa. Na matriz de confusão, observa-se que foi classificado incorretamente um número alto de registros: 465. Já o tempo de processamento foi menor do que o da MLP e maior do que o da RBF e da RF.

Diante disso, percebe-se que a classe presenteísmo foi a que apresentou, proporcionalmente, maior dificuldade na classificação nas quatro técnicas aplicadas. Já as classes normal e absenteísmo apresentaram, proporcionalmente, menor dificuldade na classificação, sendo que a MLP e a RF não mostraram erros de classificação nessas classes. Na Etapa 1, a técnica que obteve melhor desempenho, juntamente com a MLP, foi a RF. Entretanto, a RF apresentou um tempo de processamento sensivelmente menor do que o da MLP. Dessa forma, na Etapa 1, pode-se considerar a RF como a melhor escolha para o problema proposto.

4.3.2 Análise dos resultados computacionais da Etapa 2

No Quadro 17, faz-se uma comparação dos experimentos computacionais, considerando o tempo de processamento, as classes de comportamento e as matrizes de confusão da Etapa 2.

Quadro 17 – Comparação dos experimentos computacionais da Etapa 2

MLP	Classe	Presenteísta	Normal	Absenteísta	Taxa de acerto (%)	Índice Kappa	Tempo de processamento (s)
	Presenteísta	16	2	0			
Normal	0	1135	0				
Absenteísta	0	0	1250				
RBF	Presenteísta	6	9	3	97,25	0,9456	0,33
	Normal	5	1107	23			
	Absenteísta	0	26	1224			
RF	Presenteísta	16	2	0	99,91	0,9984	0,45
	Normal	0	1135	0			
	Absenteísta	0	0	1250			
ACO	Presenteísta	12	6	0	84,44	0,6885	18
	Normal	0	815	320			
	Absenteísta	0	48	1202			

Fonte: Elaborado pelo autor (2020).

a) MLP

Conforme demonstrado no Quadro 17, observa-se que, na Etapa 2, o experimento computacional, com a aplicação da técnica MLP, apresentou uma taxa de acerto de 99,91%, um índice Kappa de 0,9984, e um tempo de processamento de 179,82 segundos. Diante desses resultados, pode-se constatar que a MLP apresentou um desempenho muito bom na taxa de acerto e no índice Kappa. Na matriz de confusão, observa-se que foram classificados somente 2 registros incorretos, e que o tempo de processamento se mostrou elevado se comparado às demais técnicas.

b) RBF

Também conforme demonstrado no Quadro 17, observa-se que, na Etapa 2, o experimento computacional, com a aplicação da técnica RBF, apresentou uma taxa de acerto de 97,25%, um índice Kappa de 0,9456, e um tempo de processamento de 0,33 segundos. Sendo assim, a RBF obteve um desempenho bom na taxa de acerto e no índice Kappa. Na matriz de confusão, observa-se que foram classificados 66 registros incorretos, e que o tempo de processamento foi menor do que o da MLP, da RF e do ACO.

c) RF

Em relação à aplicação da RF, observa-se que, na Etapa 2, o experimento computacional apresentou uma taxa de acerto de 99,91%, um índice Kappa de 0,9984, e um tempo de processamento de 0,45 segundos. Com esses resultados, constata-se que a RF obteve um desempenho muito bom na taxa de acerto e no índice Kappa, igualando-se ao desempenho da MLP. Na matriz de confusão, observa-se ainda que foram classificados somente 2 registros incorretos, e que o tempo de processamento foi menor do que o da MLP e do ACO e maior do que o da RBF.

d) ACO

Já sobre o experimento computacional realizado com o ACO, observa-se que essa técnica apresentou uma taxa de acerto de 84,44%, um índice Kappa de 0,6885, e um tempo de processamento de 18 segundos. Sendo assim, o ACO obteve um desempenho inferior ao das demais técnicas na taxa de acerto e no índice Kappa. Além disso, na matriz de confusão, observa-se que foram classificados 374 registros incorretos, e que o tempo de processamento foi menor do que o da MLP e maior do que o da RBF e RF.

Com isso, percebe-se que, nas quatro técnicas aplicadas, a classe presenteísmo foi a que apresentou, proporcionalmente, maior dificuldade na classificação. As classes normal e absenteísmo, por outro lado, apresentaram, proporcionalmente, menor dificuldade na classificação, sendo que a MLP e a RF não mostraram erros de classificação, ou seja, obtiveram um melhor desempenho.

O menor tempo de processamento foi obtido com a aplicação da RBF, seguido pela RF. Dessa forma, na Etapa 2, a melhor escolha, para solucionar o problema proposto nos experimentos computacionais, continua sendo a RF. Vale ressaltar que o desempenho da RBF e do ACO melhorou em relação à Etapa 1, assim como o tempo de processamento, que diminuiu nas quatro técnicas aplicadas. Isso pode indicar que o acréscimo do atributo enriquecido com as características antropométricas e ergonômicas, por meio da rede SOM, melhorou o modelo de previsão proposto, uma vez que houve ganho no desempenho e no tempo de processamento simultaneamente.

4.3.3 Análise dos resultados computacionais da Etapa 3

No Quadro 18, faz-se uma comparação dos experimentos computacionais, considerando o tempo de processamento, as classes de comportamento e as matrizes de confusão da Etapa 3.

Quadro 18 – Comparação dos experimentos computacionais da Etapa 3

	Classe	Presenteísta	Normal	Absenteísta	Taxa de acerto (%)	Índice Kappa	Tempo de processamento (s)
MLP	Presenteísta	17	1	0	99,96	0,9992	269,61
	Normal	0	1135	0			
	Absenteísta	0	0	1250			
RBF	Presenteísta	7	8	3	96,25	0,9259	0,39
	Normal	3	1097	35			
	Absenteísta	2	39	1209			
RF	Presenteísta	16	2	0	99,91	0,9984	0,38
	Normal	0	1135	0			
	Absenteísta	0	0	1250			
ACO	Presenteísta	15	3	0	91,80	0,8368	36
	Normal	0	955	180			
	Absenteísta	0	14	1236			

Fonte: Elaborado pelo autor (2020).

a) MLP

Conforme demonstrado no Quadro 18, observa-se que, na Etapa 3, o experimento computacional, com a aplicação da técnica MLP, apresentou uma taxa de acerto de 99,96%, um índice Kappa de 0,9992, e um tempo de processamento de 269,61 segundos. Com isso, fica evidente que a MLP obteve um desempenho muito bom na taxa de acerto e no índice Kappa, sendo o melhor desempenho em comparação com as demais técnicas. Na matriz de confusão, observa-se que foi classificado somente 1 registro incorreto, e que o tempo de processamento se mostrou muito elevado se comparado às demais técnicas.

b) RBF

Ainda conforme demonstrado no Quadro 18, observa-se que, na Etapa 3, o experimento computacional, com a aplicação da técnica RBF, apresentou uma taxa de acerto de 96,25%, um índice Kappa de 0,9259, e um tempo de processamento de 0,39 segundos. Sendo assim, constata-se que a RBF obteve um desempenho bom na taxa de acerto e no índice Kappa. Entretanto, seu desempenho foi menor na taxa de acerto e no índice Kappa em relação à Etapa

2, o que pode indicar que a RBF não é tão aderente ao problema proposto, uma vez que as demais técnicas melhoraram o desempenho em comparação com a MLP e o ACO e com o tempo de processamento relativo à RF. Na matriz de confusão, observa-se que foram classificados 90 registros incorretos, e que o tempo de processamento foi menor do que o da MLP e do ACO e maior do que o da RF.

c) RF

Em relação à RF, observa-se que, na Etapa 3, o experimento computacional apresentou uma taxa de acerto de 99,91%, um índice Kappa de 0,9984, e um tempo de processamento de 0,38 segundos. Tais resultados mostram que a RF obteve um desempenho muito bom na taxa de acerto e no índice Kappa, sendo o segundo melhor desempenho, perdendo apenas para a MLP. Na matriz de confusão, observa-se que foram classificados somente 2 registros incorretos, e que o tempo de processamento foi menor do que o da MLP, da RBF e do ACO.

d) ACO

Sobre o experimento computacional realizado com o ACO, observa-se que, na Etapa 3, essa técnica apresentou uma taxa de acerto de 91,80%, um índice Kappa de 0,8368, e um tempo de processamento de 36 segundos. Dessa forma, constata-se que o ACO obteve um resultado bom na taxa de acerto e no índice Kappa, melhorando assim seu desempenho. Na matriz de confusão, observa-se que foram classificados 197 registros incorretos, e que o tempo de processamento foi menor do que o da MLP e maior do que o da RBF e RF.

Com isso, percebe-se que, nas quatro técnicas aplicadas, a classe presenteísmo foi a que apresentou, proporcionalmente, maior dificuldade na classificação. Na Etapa 3, a MLP obteve o melhor desempenho entre todas as técnicas. Já as classes normal e absenteísmo apresentaram, proporcionalmente, menor dificuldade na classificação, sendo que a MLP e a RF não mostraram erros de classificação, ou seja, obtiveram melhor desempenho na classificação dos comportamentos.

O menor tempo de processamento foi obtido com a RF, seguido pela RBF e pelo ACO. Dessa forma, na Etapa 3, a melhor escolha, para o problema proposto nos experimentos computacionais, continua sendo a RF, haja vista que a MLP apresenta o tempo de processamento elevado. Vale ressaltar que o desempenho da MLP e do ACO melhorou em relação à Etapa 2. Uma exceção foi a RBF, que piorou o desempenho na Etapa 3 em comparação com a Etapa 2.

Na Etapa 3, percebe-se ainda que o acréscimo dos atributos relacionados à antropometria e ergonomia melhorou o desempenho da MLP e do ACO, além de reduzir o tempo de processamento da RF, o que pode indicar que o modelo de previsão proposto, com o enriquecimento das bases de dados, melhora o desempenho dos classificadores e reduz o tempo de processamento. Vale ressaltar também que, na Etapa 3, com os atributos acrescidos pelo enriquecimento das bases de dados por meio da MLP 1, MLP 2 e da rede SOM, as quatro técnicas aplicadas apresentaram uma taxa de acerto acima de 90% e um índice Kappa com concordância quase perfeita, segundo a interpretação dos valores do índice Kappa proposta por Landis e Koch (1977).

4.4 CONSOLIDAÇÃO DOS RESULTADOS COMPUTACIONAIS DAS ETAPAS 1, 2 e 3

No Quadro 19, apresenta-se a consolidação dos resultados dos experimentos computacionais das Etapas 1, 2 e 3.

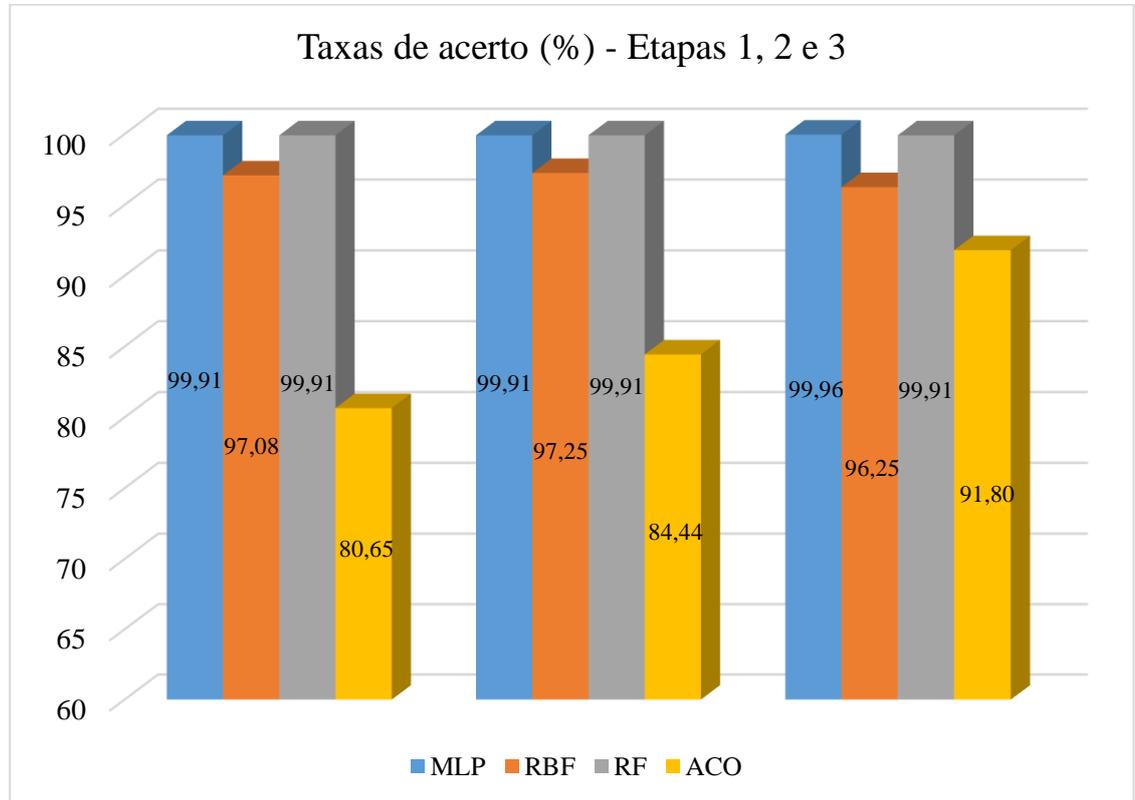
Quadro 19 – Consolidação dos experimentos computacionais das Etapas 1, 2 e 3

Etapa	Técnica	Classe	Presenteísta	Normal	Absenteísta	Taxa de acerto (%)	Índice Kappa	Tempo de processamento (s)
Etapa 1 – Base de dados Número 1 – 10 atributos	MLP	Presenteísta	16	2	0	99,91	0,9984	186,93
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	RBF	Presenteísta	6	8	4	97,08	0,9425	0,45
		Normal	6	1104	25			
		Absenteísta	7	20	1223			
	RF	Presenteísta	16	1	1	99,91	0,9984	0,48
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	ACO	Presenteísta	11	7	0	80,65	0,6127	20
		Normal	0	775	360			
		Absenteísta	0	98	1152			
Etapa 2 – Base de dados Número 2 – 11 atributos	MLP	Presenteísta	16	2	0	99,91	0,9984	179,82
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	RBF	Presenteísta	6	9	3	97,25	0,9456	0,33
		Normal	5	1107	23			
		Absenteísta	0	26	1224			
	RF	Presenteísta	16	2	0	99,91	0,9984	0,45
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	ACO	Presenteísta	12	6	0	84,44	0,6885	18
		Normal	0	815	320			
		Absenteísta	0	48	1202			
Etapa 3 – Base de dados Número 3 – 25 atributos	MLP	Presenteísta	17	1	0	99,96	0,9992	269,61
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	RBF	Presenteísta	7	8	3	96,25	0,9259	0,39
		Normal	3	1097	35			
		Absenteísta	2	39	1209			
	RF	Presenteísta	16	2	0	99,91	0,9984	0,38
		Normal	0	1135	0			
		Absenteísta	0	0	1250			
	ACO	Presenteísta	15	3	0	91,80	0,8368	36
		Normal	0	955	180			
		Absenteísta	0	14	1236			

Fonte: Elaborado pelo autor (2020).

Na Figura 32, mostra-se a taxa de acerto das quatro técnicas aplicadas na previsão do comportamento dos empregados.

Figura 32 – Taxa de acerto das quatro técnicas aplicadas para auxiliar a previsão do comportamento dos empregados



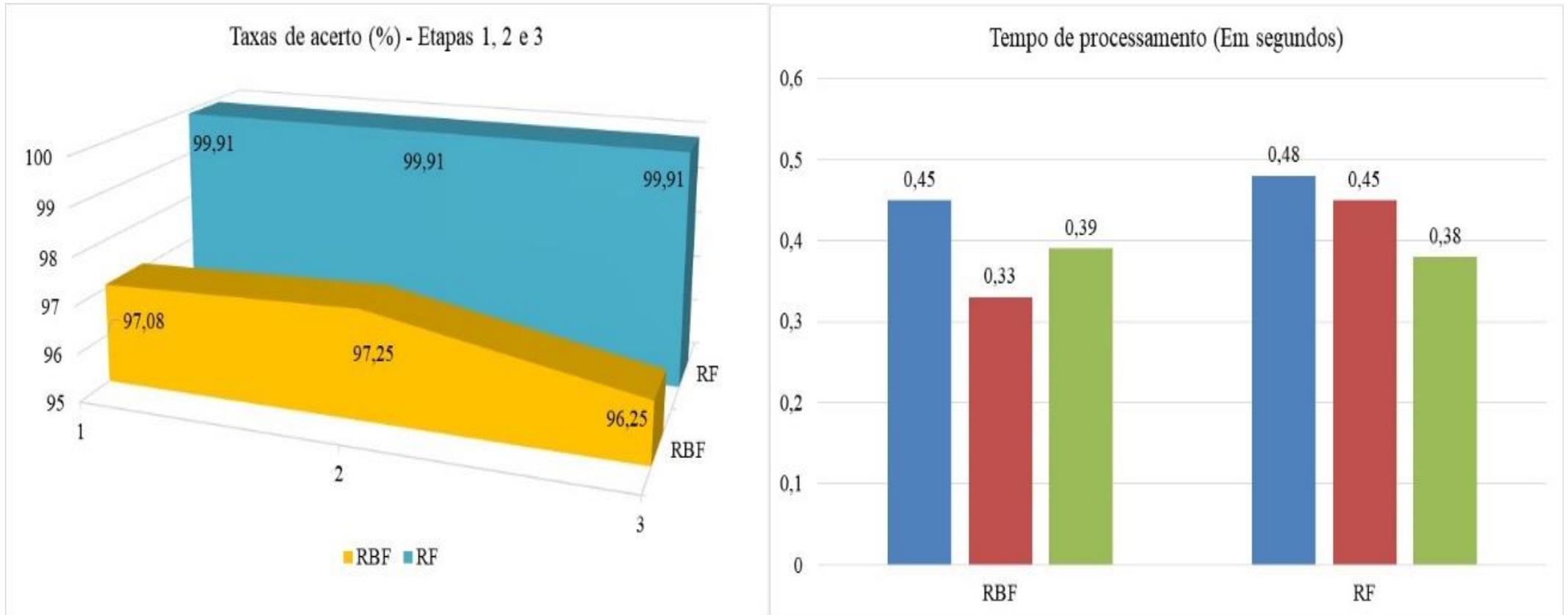
Fonte: Elaborado pelo autor (2020).

Observa-se, na Figura 32, que as quatro técnicas utilizadas apresentaram resultados, taxa de acerto, acima de 80% na Etapa 1, acima de 84% na Etapa 2 e acima de 91% na Etapa 3.

Dessa forma, entende-se que o resultado atingido por meio da utilização das técnicas foi satisfatório nas Etapas 1 e 2, exceto com o ACO. Já na Etapa 3, considera-se que o resultado obtido foi satisfatório nas quatro técnicas aplicadas.

Na Figura 33, mostra-se a taxa de acerto das duas técnicas com melhor tempo de processamento (RBF e RF) aplicadas no auxílio à previsão do comportamento dos empregados (Etapas 1, 2 e 3).

Figura 33 – Taxa de acerto e tempo de processamento das técnicas RBF e RF aplicadas para auxiliar a previsão do comportamento dos empregados



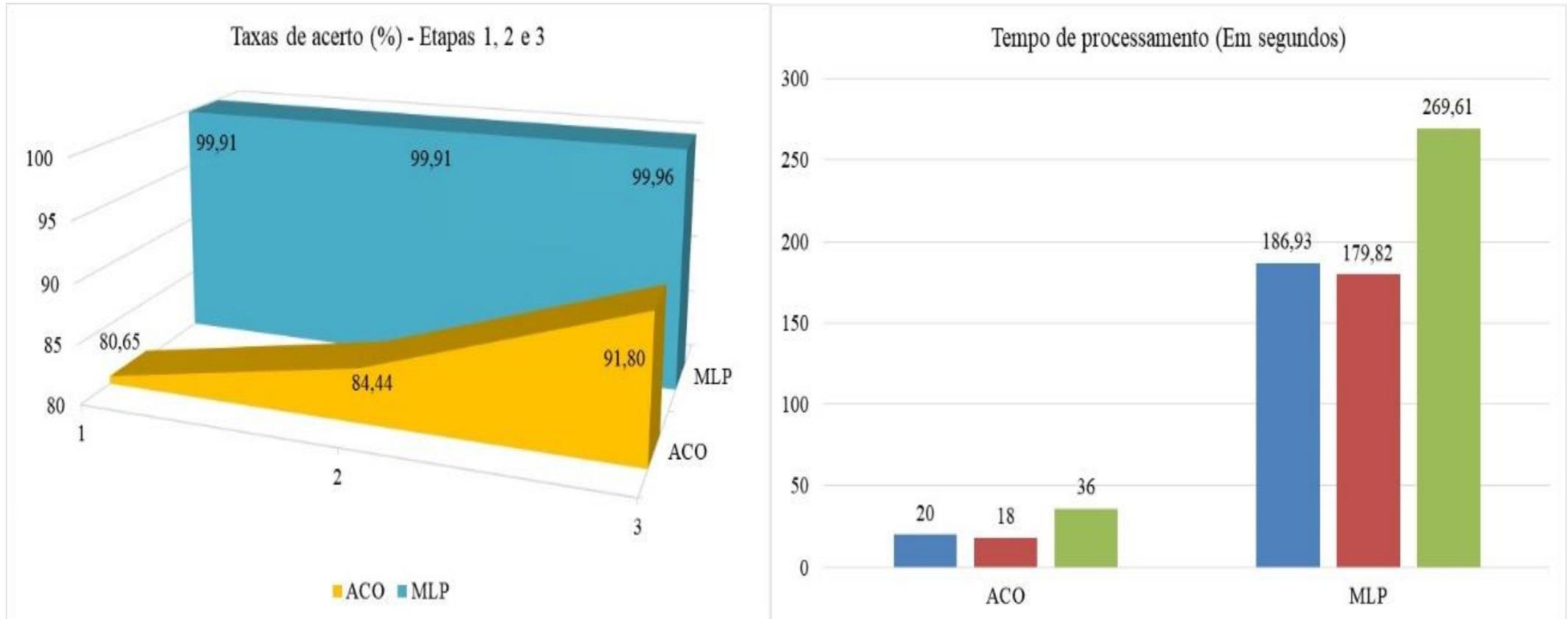
Fonte: Elaborado pelo autor (2020).

Observa-se que o desempenho apresentado pela RBF foi abaixo do desempenho apresentado pela RF. No que se refere ao tempo de processamento, as duas técnicas tiveram um resultado pequeno, menor que um segundo.

A RBF apresentou uma redução no tempo de processamento na Etapa 2 e um aumento na Etapa 3. Já a RF obteve uma redução no tempo de processamento nas Etapas 2 e 3, sendo, portanto, a única técnica que apresentou melhora no tempo de processamento nas duas etapas com o enriquecimento das bases de dados.

Na Figura 34, são demonstradas a taxa de acerto e o tempo de processamento das duas técnicas mais lentas (ACO e MLP) aplicadas na previsão do comportamento dos empregados (Etapas 1, 2 e 3).

Figura 34 – Taxa de acerto e tempo de processamento das técnicas ACO e MLP aplicadas para auxiliar a previsão do comportamento dos empregados



Fonte: Elaborado pelo autor (2020).

Observa-se que o desempenho apresentado pela MLP foi melhor do que o desempenho apresentado pelo ACO. Entretanto, o ACO apresentou melhora no desempenho nas Etapas 2 e na 3 em relação à Etapa 2. Verifica-se que o tempo de processamento das duas técnicas mais lentas foi elevado em relação às técnicas mais rápidas. Observa-se ainda que o tempo de processamento da MLP foi superior ao do ACO, que apresentou o terceiro maior tempo de processamento.

O ACO e a MLP apresentaram uma redução no tempo de processamento na Etapa 2 e um aumento na Etapa 3. É importante ressaltar que a MLP foi a técnica que mostrou o melhor desempenho entre as quatro técnicas escolhidas para a previsão do comportamento dos empregados. Verifica-se ainda que as duas técnicas apresentaram uma melhora no tempo de processamento na Etapa 2 com o enriquecimento das bases de dados.

5. CONCLUSÃO

Neste estudo, investigou-se a aplicação da mineração de dados em base de dados antropométricos, ergonômicos, absenteístas e presenteístas, para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no trabalho.

Para aplicação das técnicas de mineração de dados, foram selecionadas três bases de dados: a Número 1 com 10 atributos, a Número 2 com 11 atributos e a Número 3 com 25 atributos. Com exceção da base de dados Número 1, as bases de dados Número 2 e 3 foram enriquecidas por meio dos resultados gerados por duas redes neurais artificiais do tipo MLP e por uma rede neural artificial do tipo SOM.

Para a realização das três etapas dos experimentos computacionais, foram selecionadas quatro técnicas para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta, sendo elas: a MLP, a RBF, a RF e o ACO.

Na Etapa 1, foram obtidos os melhores desempenhos com a aplicação da MLP e da RF, seguido pela RBF e pelo ACO. É importante ressaltar que, na Etapa 1, os experimentos foram realizados com a base de dados Número 1, ou seja, a única sem o enriquecimento. A mesma ordem de desempenho foi obtida na Etapa 2. No entanto, percebeu-se uma diminuição no tempo de processamento com o acréscimo do atributo Características antropométricas e ergonômicas, o que pode indicar que, mesmo com o aumento dos dados para processamento, o atributo acrescido por meio do enriquecimento melhorou o percentual de previsão. Observou-se também que o tempo de processamento da MLP foi o mais elevado, fato que ocorreu também na Etapa 1.

Na Etapa 3, houve uma alteração na ordem de desempenho, com a MLP em primeiro lugar, seguida pela RF, pela RBF e pelo ACO, que apresentou significativo aumento na taxa de acerto. Com exceção da RBF, as demais técnicas melhoraram seu desempenho. Verificou-se que, assim como ocorreu na Etapa 2, o tempo de processamento da MLP foi bastante elevado em relação às outras técnicas.

Com isso, foi possível constatar que os resultados se mostraram significativos porque revelaram que o enriquecimento das bases de dados melhorou o desempenho das quatro técnicas aplicadas quando se considerou a taxa de acerto, que passou dos 90%, na Etapa 3.

Assim, pode-se responder afirmativamente para a questão de pesquisa deste trabalho: A mineração de dados aplicada em base de dados antropométricos, ergonômicos, absenteístas e presenteístas pode auxiliar na previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no trabalho?

Pode-se considerar também que o objetivo geral deste trabalho foi atingido ao investigar a aplicação da mineração de dados em base de dados antropométricos, ergonômicos, absenteístas e presenteístas para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta dos empregados no trabalho. Como resultado desta investigação, pode-se considerar ainda que técnicas distintas auxiliaram na previsão à medida que o enriquecimento por meio de novos atributos foi acrescentado.

Em relação ao tempo de processamento e desempenho, a RF se destaca como sendo a técnica recomendada entre as demais para auxiliar na previsão dos comportamentos presenteísta, normal e absenteísta, no trabalho.

Uma questão que pode ser discutida na continuidade deste trabalho é a análise do desempenho, com foco no tempo de processamento, para entender por que a MLP apresentou valores maiores do que os das outras três técnicas.

Como limitação deste trabalho, considera-se a escassez de aplicação de técnicas inteligentes no estudo da ergonomia e antropometria relacionadas ao absenteísmo e ao presenteísmo, conforme comprovado na revisão bibliográfica realizada, uma vez que não foi possível fazer uma comparação dos resultados dos experimentos computacionais obtidos com outros trabalhos da literatura. Considera-se também como limitação, a escassez de bases de dados públicas sobre absenteísmo e presenteísmo para aplicar técnicas inteligentes.

A contribuição para a academia encontra-se na aplicação de técnicas inteligentes, no auxílio à previsão de comportamentos de empregados, por meio do enriquecimento de bases de dados. Considera-se esta investigação pioneira no campo de estudo, o que abre caminho para novas pesquisas. Além disso, o processo de enriquecimento serve como roteiro para estudos em outras áreas.

Por fim, o enriquecimento da base de dados pode proporcionar um campo promissor para estudos em que as bases de dados disponíveis não contemplam todos os dados necessários para um determinado estudo. Dessa forma, acredita-se que o trabalho pode subsidiar novas descobertas no campo do enriquecimento de base de dados em problemas semelhantes em outras áreas do conhecimento.

Para as organizações, tem-se a oportunidade de aplicar técnicas inteligentes seguindo uma tendência mundial de aplicação dessas técnicas em diversos problemas do mundo real. Ademais, a previsão do comportamento dos empregados no trabalho permite a elaboração prévia de estratégias corporativas para reduzir os comportamentos absenteísta e presenteísta, além de permitir que os empregados com comportamento normal sejam de alguma forma valorizados, criando um ambiente favorável para a redução dos comportamentos absenteísta e

presenteísta. Importante ressaltar que a redução do absenteísmo e presenteísmo implica diretamente na redução de custos empresariais e na melhoria da qualidade do trabalho oferecido pela empresa.

Para os empregados, a contribuição principal está relacionada com o ganho na qualidade de vida, uma vez que ambientes de trabalho com altos índices de absenteísmo e presenteísmo criam grupos desmotivados, estressados e depressivos, culminando em empresas pouco criativas e pouco inovadoras, já que o prazer do trabalho por um objetivo comum deixa de existir em grupos sem nenhuma motivação. Assim sendo, a contribuição do trabalho para os empregados é fundamental para criar um ambiente de trabalho saudável e criativo e inovador.

Empregados satisfeitos naturalmente são mais produtivos e tendem a disseminar um ambiente mais agradável e cooperativo. Dessa forma, entende-se que o incentivo ao trabalho produtivo, inverso do presenteísmo, e à assiduidade, inverso do absenteísmo, traz benefícios importantes para o empregado, já que a integração com o ambiente de trabalho acaba trazendo uma satisfação pessoal e uma sensação prazerosa, criando grupos mais coesos, inovadores e, principalmente, com mais saúde, o que pode-se chamar de empregados motivados ou simplesmente felizes com o trabalho, sentimento este que pode refletir na vida pessoal.

Para a sociedade, a contribuição está relacionada com a possível redução dos afastamentos dos postos de trabalho e do custo previdenciário, que acaba rateado pela sociedade. A realização de atividades laborais de maneira errada, sem a observação das práticas corretas de ergonomia e sem um estudo antropométrico, resulta em altos índices de absenteísmo ou presenteísmo e em altos índices de aposentadorias por invalidez. Tais problemas podem afetar o desempenho da função social, como perda do condicionamento físico, deterioração da saúde geral, diminuição da participação em atividades sociais, estresse familiar, diminuição do contato com a comunidade, diminuição do poder aquisitivo, irritação, ansiedade e depressão, o que pode causar um aumento significativo nos auxílios-doença concedidos pelo Instituto Nacional do Seguro Social, onerando o Estado e, por conseguinte, a sociedade como um todo.

Como dar continuidade à pesquisa, pretende-se ampliar o enriquecimento das bases de dados com atributos relacionados à alteração climática ao longo do tempo, como mudanças de temperatura, precipitação e outros fenômenos do clima, uma vez que estudos recentes apontam que o estresse térmico afeta a produtividade do trabalho.

Dessa forma, com a continuidade da pesquisa, espera-se evidenciar a relação do absenteísmo e do presenteísmo com o estresse térmico, uma vez que esse fator afetará a maioria das regiões do mundo, e que os países em desenvolvimento serão mais afetados pela alteração climática do que os países desenvolvidos.

Por meio da realização deste estudo, percebeu-se também que outros fatores relacionados ao absenteísmo estão sendo incorporados na literatura, como o aquecimento global e a escassez de água no planeta. Com isso, tais temas se mostram como um campo de pesquisa promissor para continuidade da pesquisa, uma vez que é possível enriquecer a base de dados deste estudo com dados históricos sobre a alteração climática nas regiões.

Espera-se também, como continuidade da pesquisa, aplicar técnicas combinadas de inteligência artificial para auxiliar a previsão dos comportamentos presenteísta, normal e absenteísta, utilizando os dados da alteração climática ao longo do tempo no período de janeiro de 2008 a dezembro de 2017, ou seja, período contemplado pela base de dados do presente trabalho.

Por fim, na revisão bibliográfica realizada, foi possível verificar que, nos últimos anos, a aplicação de técnicas da inteligência artificial no estudo do absenteísmo aumentou significativamente, o que demonstra interesse crescente pelo assunto (Capítulo 3, seção 3.2 Tabela 1). Isso evidencia a relevância e contribuição deste estudo para abrir caminho para novas pesquisas sobre o tema.

5.1 MEMORIAL DE PUBLICAÇÕES

- Artigos completos publicados em periódicos

Ferreira, R. P.; Martiniano, A.; Napolitano, D. M. R.; Farias, e. B. P.; Sassi, R. J. Artificial neural network and their application in the prediction of absenteeism at work. *International Journal of Recent Scientific Research*, v. 9, p. 23332-23334, 2018.

- Trabalhos completos publicados em anais de congressos

Martiniano, A.; Ferreira, R. P.; Sassi, R. J. Mineração de dados aplicada a um estudo sobre ergonomia em uma empresa de logística: conhecimento e prática. In: *Simpósio de Excelência em Gestão e Tecnologia - SEGeT, 2018, Resende - RJ. A indústria 4.0 e o uso de tecnologias digitais. Resende/Rio de Janeiro: SEGeT. p. 1-16, 2018.*

- Capítulo de livro:

Martiniano, A.; Ferreira, R. P.; Farias, E. B. P.; Gatto, D. D. O.; Sassi, R. J. Mineração de dados aplicada ao Absenteísmo e a Ergonomia. In: *Editora Poisson. (Org.). Tópicos em*

Administração - Capítulo 17 - Mineração de dados aplicada ao Absenteísmo e a Ergonomia. 1ª ed. Belo Horizonte: Poisson, v.8, p. 204-217, 2018.

5.2 BASES DE DADOS CEDIDAS PARA UCI *MACHINE LEARNING REPOSITORY*

- Base de Dados cedida para UCI *Machine Learning Repository* do Centro de Aprendizado de Máquina e Sistemas Inteligentes da Universidade da Califórnia Irvine:

Martiniano, A.; Ferreira, R. P.; Sassi, R. J. Absenteeism at work Data set - Base de Dados cedida para UCI *Machine Learning Repository* (<http://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>). 2018.

A base de dados foi cedida para UCI *Machine Learning Repository* com o objetivo de popularizar os dados das pesquisas realizadas na Universidade Nove de Julho e para que outros estudantes e pesquisadores possam reproduzir os experimentos.

- Vídeo aulas no youtube

- *K-Means Clustering: Absenteeism at work Data Set*, 2019. Disponível em: <https://www.youtube.com/watch?v=bW8mzMO2fVE>

- *Data Mining Clustering Meetode K-Means dan Hierarchial Dataset Absenteeism at Work*, 2019. Disponível em: <https://www.youtube.com/watch?v=l7yldvNu-FE>

- Citações da base *Absenteeism at work Data set* no Google Acadêmico

- DOGRUYOL, K.; SEKEROGLU, B. Absenteeism Prediction: A Comparative Study Using Machine Learning Models. In: International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions. Springer, Cham, v.1095, p. 728-734, 2020.

- TEWARI, K.; VANDITA, S.; JAIN, S. Predictive Analysis of Absenteeism in MNCS Using Machine Learning Algorithm. In: Proceedings of ICRIC 2019. Springer, Cham, 2020. p. 3-14, 2020.

- ARAUJO, V. S.; REZENDE, T. S.; GUIMARÃES, A. J.; ARAUJO, V. J. S.; CAMPOS SOUZA, P. V. A hybrid approach of intelligent systems to help predict absenteeism at work in companies. *SN Applied Sciences*, v.1, n.6, p.536, 2019.
- ASIRI, A.; ABDULLAH, M. Employees Absenteeism Factors Based on Data Analysis and Classification. *Special Issue in Communication and Information Technology*, v.12, p.119-127, 2019.
- PATEL, S.; PATEL, A. Performance Analysis and Evaluation of Clustering Algorithms. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, v.8, p.179-183, 2019.
- PRIYANKA, D.; NAYAK, J. Empirical Analysis of Absenteeism at Work Place Using Machine Learning. In: *International Conference on Application of Robotics in Industry using Advanced Mechanisms*. Springer, Cham, p.150-160, 2019.
- WAHID, Z.; SATTER, A. K. M.; AL IMRAN, A.; BHUIYAN, T. Predicting Absenteeism at Work Using Tree-Based Learners. In *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, p.7-11, 2019.
- GAYATHRI, T. Data mining of Absentee data to increase productivity. *International Journal of Engineering and Techniques*. v.4, n.3, p.478-480, 2018.
- PAAKAALA, A. V. R.; MACHA, S. S.; MUDIGONDA, K. S. Evaluation of Clustering Algorithms on Absenteeism at Work Dataset. *IJSRD - International Journal for Scientific Research & Development*. v.6, Issue 06, p.337-342, 2018.

- Citações da base *Absenteeism at work Data set* em outra base

- LARSSON, C. Case Representation Methodology for a Scalable Case-Based Reasoning. Malardalen University. School of Innovation, Design and Engineering - Vasteras, Sweden, 2018.

5.3 REGISTRO DE PROGRAMA DE COMPUTADOR ORIUNDO DA PESQUISA

Ferreira, R. P.; Martiniano, A.; Sassi, R. J. Software Tool for Absenteeism Prediction. 2018. Patente: Programa de Computador. Número do registro: BR512018001297-3, data de registro: 14/01/2018, título: "*Software Tool for Absenteeism Prediction*", Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial. (Anexo A).

REFERÊNCIAS BIBLIOGRÁFICAS

- ABERGO, Associação brasileira de Ergonomia. Disponível em: < http://www.abergo.org.br/internas.php?pg=o_que_e_ergonomia>. Acesso em: 15 de jan. 2020.
- ADDAE, H. M.; JOHNS, G.; BOIES, K. The legitimacy of absenteeism from work: a nine nation exploratory study. **Cross Cultural Management: An International Journal**, v. 20, n. 3, p. 402-428, 2013.
- AL-BEHADILI, H. N. K.; KU-MAHAMUD, K. R.; SAGBAN, R. Ant colony optimization algorithm for rule-based classification: Issues and potential solutions. **Journal of Theoretical and Applied Information Technology**, v.96, n.21, 2018.
- ALSANAD, A. Forecasting Daily Demand of Orders Using Random Forest Classifier. **International Journal of Computer Science and Network Security - IJCSNS**, v.18, n.4, p.79-83, 2018.
- ANDRADE, C. G. A utilização do algoritmo de árvore de decisão C4.5 para a gestão do absenteísmo organizacional. *Postal Brasil – Revista técnico científica dos Correios*, v.2, p.61-70, 2018.
- APPOLINÁRIO, F. Dicionário de Metodologia Científica. 2. ed. São Paulo: Atlas, 2011.
- ARAÚJO, V. S.; REZENDE, T. S.; GUIMARÃES, A. J.; ARAÚJO, V. J. S.; DE CAMPOS S. P. V. A hybrid approach of intelligent systems to help predict absenteeism at work in companies. **SN Applied Sciences**, v.1, n.6, p.536, 2019.
- ASIRI, A.; ABDULLAH, M. Employees Absenteeism Factors Based on Data Analysis and Classification. **Special Issue in Communication and Information Technology**, v.12, p.119-127, 2019.
- ASSUNÇÃO, A. Á.; ABREU, M. N. S. Fatores associados a distúrbios osteomusculares relacionados ao trabalho autorreferidos em adultos brasileiros. *Revista de Saúde Pública*, v.51, p.1-12, 2017.
- BACK, T. Adaptative Business Intelligence based on evolution strategies: some application examples of self-adaptative software. **Information Sciences**, v.148, n.1-4, p.113-121, 2002.
- BAE, Y. H. Relationships between presenteeism and work-related musculoskeletal disorders among physical therapists in the Republic of Korea. **International Journal of Occupational Safety and Ergonomics**, v. 24, n. 3, p. 487-492, 2018.
- BANERJEE, M.; CAPOZZOLI, M.; MCSWEENEY, L.; SINHA, D. Beyond kappa: A review of interrater agreement measures. **Canadian Journal of Statistics**, v.27, n.1, p. 3-23, 1999.
- BATIZ, E. C.; VERGARA, L. G.; LICEA, O. E. A. Análise comparativa entre métodos de carregamento de cargas e análise postural de auxiliares de enfermagem. **Production**, v.22, n.2, p.270-283, 2012.
- BERNSTROM, V. H. The relationship between three stages of job change and long-term sickness absence. **Social Science & Medicine**. v.98, p.239–246, 2013.
- BIGUS, J. P. Data Mining with Neural Network: Solving Business Problems from Applications Development to Decision Support. **Mcgraw-Hill**, 1996.

- 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.
- BLUM, A.; MERINO, E. A. D.; MERINO, G. S. A. D. **Método visual para revisão sistemática em Design com base em conceitos da Mineração de Dados**. DAPesquisa, v.11, n.16, p.124-139, 2016.
- BÖCKERMAN, P.; LAUKKANEN, E. Predictors of sickness absence and presenteeism: does the pattern differ by a respondent's health? **Journal of Occupational and Environmental Medicine**, v.52, n.3, p.332-335, 2010.
- BÖCKERMAN, P.; LAUKKANEN, E. Presenteeism in Finland: Determinants by gender and the sector of economy. **Ege Academic Review**, v.9, n.3, p.1007-1016, 2009.
- BONABEAU, E.; DORIGO, M.; THERAULAZ, G. *Swarm Intelligence: from natural to artificial systems*. Oxford University Press, Inc., 1999.
- BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**, 2ª ed. Rio de Janeiro: LTC, 2011
- BRAGANÇA, S.; AREZES, P.; CARVALHO, M.; ASHDOWN, S. P. Current state of the art and enduring issues in anthropometric data collection. *Dyna*. v.83, n.197, p.22-30, 2016.
- BRAHAMI, M.; ATMANI, B.; MATTA, N. Dynamic knowledge mapping guided by data mining: application on healthcare. **Journal of Information Processing Systems**, v.9, n.1, p.1-30, 2013.
- BRATSBERG, B.; FEVANG, E.; RØED, K. Job loss and disability insurance. **Labour Economics**, v.24, p.137–150, 2013.
- BREARD, G. T. Evaluating self-organizing map quality measures as convergence criteria. Master thesis, University of Rhode Island, 2017.
- BREIMAN, L. Random forests. *Machine learning*. **Springer**, v.45, n.1, p.5–32, 2001.
- BREIMAN, L.; CUTLER, A. *An introduction to random forest for beginners*. 1. ed. Califórnia, Estados Unidos: **Salford Systems**, v.71 p.24, 2014.
- BRITT, T. W.; JEX, S. M. *Thriving under stress: Harnessing demands in the workplace*. Oxford **University Press**, USA, 2015.
- BUSTILLOS, A. S.; VARGAS III, K. G.; GOMERO-CUADRA, R. Work productivity among adults with varied Body Mass Index: Results from a Canadian population-based survey. **Journal of Epidemiology and Global Health**, v.5, n.2, p.191-199, 2015.
- CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A. *Discovering Data Mining: from concept to implementation*. New Jersey: Prentice Hall, 1998.
- CALVO-DAHLBORG, M.; BROWN, S. G. R. Hume-Rothery for HEA classification and self-organizing map for phases and properties prediction. **Journal of Alloys and Compounds**, v. 724, p.353-364, 2017.
- CARDOSO, D. S.; LEITE FILHO, H. P.; SOUTO, R. **Técnicas de extração de conhecimento por meio de dados faltantes e mineração de dados sobre as vítimas do césio-137: Uso De**

Imputação Múltipla Para Obtenção De Modelos Acurados. Multi-Science Journal, v.1, n.7, p.79-91, 2018.

CARDOSO, O. N. P.; MACHADO, R. T. M. Knowledge management using data mining: a case study of the Federal University of Lavras. **Revista de Administração Pública**, v.42, n.3, p.495-528, 2008.

CARVALHO, D. R.; MOSER, A. D.; DA SILVA, V. A.; DALLAGASSA, M. R. **Mineração de dados aplicada à fisioterapia.** Fisioterapia em Movimento, v.25, n.3, p.595-605, 2012.

CHANDRASEKAR, K. Workplace environment and its impact on organisational performance in public sector organisations. **International journal of enterprise computing and business systems**, v.1, n.1, p.1-19, 2011.

COLLINS, A. M.; CARTWRIGHT, S.; COWLISHAW, S. Sickness presenteeism and sickness absence over time: A UK employee perspective. **Work & Stress**, v. 32, n. 1, p. 68-83, 2018.

COSTA, L.; RAMOS, R.; MORO, S. Anticipating Next Public Administration Employees Absence Duration. **Public administration issues**, v.6, p.23-40, 2019.

DA SILVA, J. D. S. C.; FELTRIN, J. O.; SCHWALM, M. T.; CORREA, S. M. **Principais fatores de Absenteísmo na enfermagem de um Hospital de Médio Porte do Sul Catarinense.** Revista Interdisciplinar de Estudos em Saúde, v.7, n.1, p.319-334, 2018.

DE AVELAR, C. F. P.; ROCHA, T. A. H.; CRUZ, F. J. S. **Mineração de dados: uma revisão da literatura em Administração.** Revista Vianna Sapiens, v.8, n.2, p.30-54, 2017.

DEFANI, J. C. **Avaliação do perfil antropométrico e análise dinamométrica dos trabalhadores da agroindústria do setor de frigoríficos e abatedouros: O Caso da Perdígão-Carambeí,** 143 f. Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2007.

DE OLIVEIRA; T.; MOREIRA, R. S.; DE LIMA, R. A. F. Absenteeism prediction in call center using machine learning algorithms. **In World Conference on Information Systems and Technologies.** Springer, Cham, p.958-968, 2019.

DE REUVER, R.; VAN DE VOORDE, K.; KILROY, S. When do bundles of high performance work systems reduce employee absenteeism? The moderating role of workload. **The International Journal of Human Resource Management**, v. 30, p.1-21, 2019.

DIONNE, C. E.; DUNN, K.M., CROFT, P.R., NACHEMSON, A.L., BUCHBINDER, R.; WALKER, B.F. A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine*. v.33, n.1, p.95-103, 2008.

DOGRUYOL, K.; SEKEROGLU, B. Absenteeism Prediction: A Comparative Study Using Machine Learning Models. In: **International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions.** Springer, Cham, v.1095, p. 728-734, 2020.

DORIGO, M.; STÜTZLE, T. Ant Colony Optimization. **Bradford Book**, 2004.

DRAKOPOULOS, S. A.; GRIMANI, K. Injury-related absenteeism and job satisfaction: Insights from Greek and UK data. **The International Journal of Human Resource Management**, v.24, n. 18, p.3496-3511, 2013.

DUL, J.; WEERDMEEESTER, B. **Ergonomia prática.** São Paulo: Edgar Blücher, 2004.

- EVANS-LACKO, S.; KNAPP, M. Global patterns of workplace productivity for people with depression: Absenteeism and presenteeism costs across eight diverse countries. **Social Psychiatry and Psychiatric Epidemiology**, v.51, p.1525–1537, 2016.
- FARID, D. M.; ZHANG, L.; RAHMAN, C. M.; HOSSAIN, M. A.; STRACHAN, R. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. **Expert Systems with Applications**, v.41, n.4, p.1937-1946, 2014.
- FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMITH, P. *The KDD process for extracting useful knowledge from volumes of data*. **Communications of the ACM**, v.39, p.27-34, 1996.
- FENG, M.; ZHENG, J.; REN, J.; HUSSAIN, A.; LI, X.; XI, Y.; LIU, Q. Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data. **IEEE Access**, v.7, p.106111-106123, 2019.
- FERREIRA, M. C. **Ergonomia da Atividade aplicada à Qualidade de Vida no Trabalho: lugar, importância e contribuição da Análise Ergonômica do Trabalho (AET)**. Revista Brasileira de Saúde Ocupacional, v.40, n.131, p.18-29, 2015.
- FERREIRA, R. P.; MARTINIANO, A. M.; NAPOLITANO, D.; FARIAS, E. B. P.; SASSI, R. J. Artificial neural network and their application in the prediction of Absenteeism at work. **International Journal of Recent Scientific Research Research**, v.9, p.23332-23334, 1ª ed., 2018.
- FERREIRA, R. P.; MARTINIANO, A. S.; SASSI, R. J. Um estudo usando *data mining* com o algoritmo density based scan algorithm with noise (DBSCAN) na clusterização de perfis de empregados ausenteístas e presenteístas. In: **13º International conference on information systems & technology management**. CONTECSI. São Paulo: CONTECSI, v. XIII. p. 1-22, 2016a.
- FERREIRA, R.; MARTINIANO, A.; SASSI, R. J. **Dashboard Inteligente para apoio à tomada de decisão em empresa de courier**. Revista Gestão & Tecnologia, v.16, n.2, p.39-72, 2016b.
- FERREIRA, R. P.; MARTINIANO, A.; SASSI, R. J. Software Tool for Absenteeism Prediction. 2018. Patente: Programa de Computador. Número do registro: BR512018001297-3, data de registro: 14/01/2018, título: "Software Tool for Absenteeism Prediction", Instituição de registro: INPI - **Instituto Nacional da Propriedade Industrial**, 2018.
- FERREIRA, M. S.; NAVEGA, M. T. Effects of a guidance program to adults with low back pain. **Acta Ortopédica Brasileira**, v.18, n.3, p.127-131, 2010.
- FERREIRA, R. P. **Inteligência Computacional na Previsão do Absenteísmo e Identificação de Tendências Absenteístas**. 2019. 200f. Tese (Doutorado em Informática e Gestão do Conhecimento). Universidade Nove de Julho. São Paulo, 2019.
- FERRO, D.; ZACHARIAS, F. C. M.; FABRIZ, L. A.; SCHONHOLZER, T. E.; VALENTE, S. H.; BARBOSA, S. M.; PINTO, I. C. **Absenteísmo na equipe de enfermagem em serviços de emergência: implicações na assistência**. Acta Paulista de Enfermagem, v.31, n.4, p.399-408, 2018.
- FLEISS, J. L., LEVIN, B., &PAIK, M. C. The Measurement of Interrater Agreement. **Statistical Methods for Rates and Proportions, Third Edition**, p. 598-626, Willey, 2003.
- FLORES, L. I.; VILELA, L. O.; BORELLI, L. M.; JÚNIOR, E. G.; CAMARGO, M. L. **O absenteísmo enquanto indicador para o processo de gestão de pessoas nas organizações e de atenção à saúde do trabalhador**. Rev. Laborativa, v.5, n.2, p.47-65, 2016.

- FONSECA, J.; SANTOS, C. A. Anatomia Aplicada à Clínica: **Antropometria na Avaliação Nutricional de 367 Adultos Submetidos a Gastrostomia Endoscópica**. Acta Médica portuguesa, v.26, n.3, 2013.
- FONSECA, S. O.; NAMEN, A. A. *Data Mining on Inep Databases: An Initial Analysis Aiming to Improve Brazilian Educational System*. **Educ. rev.** v.32 n.1 Belo Horizonte Jan./Mar. 2016.
- FOTOVATIKHAH, F.; HERRERA, M.; SHAMSHIRBAND, S.; CHAU, K. W.; ARDABILI, S. F.; PIRAN, M. J. Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. **Engineering Applications of Computational Fluid Mechanics**, v.12, n.1, p.411-437, 2018.
- FRANCO, T.; DRUCK, G.; SELIGMANN-SILVA, E. **As novas relações de trabalho, o desgaste mental do trabalhador e os transtornos mentais no trabalho precarizado**. Revista Brasileira de Saúde Ocupacional, v.35, n.122, p.229-248, 2010.
- FREITAS, A.; PARPINELLI, R.; LOPES, H. Ant colony algorithms for data classification. **To appear in Encyclopedia of Info. Sci. & Tech.** 2 ed., 2008.
- FRITZSCHE, L.; WEGGE, J.; SCHMAUDER, M.; KLIEGEL, M.; SCHMIDT, K. H. Good ergonomics and team diversity reduce absenteeism and errors in car manufacturing. **Ergonomics**, v.57, n.2, p.148-161, 2014.
- GARRIDO, G.; BORGES, M. K.; BORGES, R. S.; SILVEIRA, M. A. **Métricas do presenteísmo e suas relações com a cooperação: Uma evidência empírica**. Revista de Administração Mackenzie, v.20, n.2, p.-1-30, 2019.
- GARRIDO, G.; MENDONÇA, A. V.; LOPES, K. M. O. L.; SILVEIRA, M. A. **Presenteísmo: Causas e consequências de um mal subterrâneo**. Revista de Ciências da Administração, v. 19, n.48, p.54-67, 2017.
- GAYATHRI, T. Data mining of Absentee data to increase productivity. **International Journal of Engineering and Techniques**, v.4, n.3, p.478-480, 2018.
- GAUDART, C. Activity, time and itineraries: for the integration of multiple times in the ergonomic analysis of work. **Le travail humain**, v. 79, n. 3, p. 209-232, 2016.
- GOLDSCHMIDT, R. R.; PASSOS, E.; BEZERRA E. **Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2ª ed. Campus: Rio de Janeiro, 2015.
- GUALHANO, M. A.; SALLES, S. A. F.; DA HORA, H. R. M. **Mineração de dados das fichas da Avaliação Quadrienal da Capes dos Programas da área Interdisciplinar: Engenharia, Tecnologia e Gestão**. Revista Meta: Avaliação, v.10, n.29, p.417-442, 2018.
- GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 6ª. ed. São Paulo: Atlas, 2018.
- GROSS, H. P.; THALER, J.; WINTER, V. Integrating Public Service Motivation in the Job-Demands-Resources Model: An Empirical Analysis to Explain Employees' Performance, Absenteeism, and Presenteeism. **International Public Management Journal**, v.22, n.1, p. 176-206, 2019.
- HALBESLEBEN, J. R. B.; WHITMAN, M. V.; CRAWFORD, W. S. A dialectical theory of the decision to go to work: Bringing together absenteeism and presenteeism. **Human Resource Management Review**, v.24, p.177-192, 2014.
- HAEFFNER, R.; KALINKE, L. P.; FELLI, V. E. A.; MANTOVANI, M. D. F.; CONSONNI, D.; SARQUIS, L. M. M. **Absenteísmo por distúrbios musculoesqueléticos em**

trabalhadores do Brasil: milhares de dias de trabalho perdidos. Revista Brasileira de Epidemiologia, v.21, p. e180003, 2018.

HAN, J.; KAMBER, M.; PEI, J. Data mining concepts and techniques. Estados Unidos: **Morgan Kaufmann and Elsevier**. 3. ed., 2012.

HANSEN, C. D.; ANDERSEN, J. H. Going ill to work – What personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? **Social Science and Medicine**, v.67, n.6, p.956–964, 2008.

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

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

IBRAHIM, I. A.; KHATIB, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. **Energy Conversion and Management**, v.138, p.413-425, 2017.

IIDA, I.; BUARQUE, L. **Ergonomia: projeto e produção**. Editora Blucher, 2016.

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, v.61, n.2, Brasília, 2008.

JACKSON FILHO, J. M. J.; LIMA, F. P.A. Ergonomic Work Analysis in Brazil: a successful technological transfer? **Rev. Brasileira de Saúde ocupacional**, São Paulo, v.40, n.131, p.12-17, 2015.

JANTAN, H.; HAMDAN, A.R.; OTHMAN, Z.A. Towards applying Data Mining Techniques for Talent Mangement. **In International Conference on Computer Engineering and Applications**, (2). IACSIT, 2009.

JOHANSEN, V.; ARONSSON, G.; MARKLUND, S. Positive and negative reasons for sickness presenteeism in Norway and Sweden: a cross-sectional survey. **BMJ open**, v.4, n.2, p. e004123, 2014.

JOHNS, G. Absenteeism and presenteeism: Not at work or not working well. **The Sage handbook of organizational behavior**, v.1, p.160-177, London, 2008.

JOHNS, G. How methodological diversity has improved our understanding of absenteeism from work. **Human resource management review**, v.13, n.2, p.157-184, 2003.

JOHNS, G. Presenteeism in the workplace: A review and research agenda. **Journal of Organizational Behavior**, v.31, n.4, p.519-542, 2010.

JOHNS, G. The essential impact of context on organizational behavior. **Academy of management review**, v.31, n.2, p.386-408, 2006.

KARANIKA-MURRAY, M.; PONTES, H. M.; GRIFFITHS, M. D.; BIRON, C. Sickness presenteeism determines job satisfaction via affective-motivational states. **Social Science & Medicine**, v.139, p.100-106, 2015.

KARWOWSKI, W. The discipline of ergonomics and human factors. **Handbook of human factors and ergonomics**, v.3, p.3-31, 2006.

KAVAKIOTIS, I.; TSAVE, O.; SALIFOGLU, A.; MAGLAVERAS, N.; VLAHAVAS, I.; CHOUVARDA, I. Machine learning and data mining methods in diabetes research. **Computational and structural biotechnology journal**, v.15, p.104-116, 2017.

- KESAVARAJ, G.; SUKUMARAN, S. A study on classification techniques in data mining. In: 2013 Fourth International Conference on Computing, **Communications and Networking Technologies** (ICCCNT). IEEE, p.1-7, 2013.
- KIVILUOTO, K. Topology Preservation in Self-Organizing Maps. In Proc. **International Conf. Neural Networks**, Washington, DC. p.294–299, 1996.
- KLEIN, L. L.; PEREIRA, B. A.; LEMOS, R. B. **Qualidade de vida no trabalho: parâmetros e avaliação no serviço público**. RAM. Revista de Administração Mackenzie, v.20, n.3., 2019.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. KOHONEN, Teuvo. Self-organized formation of topologically correct feature maps. **Biological cybernetics**, v.43, n.1, p.59-69, 1982.
- KOHONEN, T. Self-Organizing Maps. New York. **Springer**. 3ª ed., 2001.
- KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, v.37, p. 52-65, 2013.
- KOVÁCS, Z. L. **Redes Neurais Artificiais - Fundamentos e Aplicações**. Editora Livraria da Física. 4ª ed. São Paulo, 2006.
- KROEMER, K. H. E.; GRANDEJEAN, E. **Manual de Ergonomia: Adaptando o Trabalho ao Homem**. Porto Alegre: Bookman editora, 2005.
- KUMAR R.; SRIVASTAVA S.; GUPTA J. Modeling and adaptive control of nonlinear dynamical systems using radial basis function network. **Soft Comput**, p.1-17, 2017.
- LANDIS, J., KOCH, G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p.159–174, 1977.
- LARANJEIRA, C. A. **O Contexto Organizacional e a Experiência de Estress: uma Perspectiva Integrativa**. Revista *salud pública*, v.11, n.1, p.123-133, 2009.
- LARDON, A.; DUBOIS, J. D.; CANTIN, V.; PICHÉ, M.; DESCARREAUX, M. Predictors of disability and absenteeism in workers with non-specific low back pain: A longitudinal 15-month study. **Applied ergonomics**, v.68, p.176-185, 2018.
- LAROSE, D. T.; LAROSE, C. D. Discovering knowledge in data: an introduction to data mining. **John Wiley & Sons. New Jersey**, 2ª ed., 2014.
- LEANDRO, G.; PAROLIM, S. C.; MORO, C. M. C.; CARVALHO, D. R. **Mineração de dados na avaliação de óbitos após cirurgia de amputação**. *Jornal vascular brasileiro*, v.17, n.1, p.10-18, 2018.
- LINOFF, G. S., BERRY, M. J. Data mining techniques: for marketing, sales, and customer relationship management. **John Wiley & Sons**, 3ª ed., 2011.
- LOHAUS, D.; HABERMANN, W. Presenteeism: A review and research directions, v.29, n.1, p.43-58, 2019.
- LOPES, K. M. O.; GARRIDO, G.; MENDONÇA, A. V.; SILVEIRA, M. A. **Uma Abordagem Fuzzy para o estudo do Presenteísmo**. *Revista Pensamento Contemporâneo em Administração*, v.11, n.2, p.136-150, 2017.
- MACGREGOR, J. N.; CUNNINGHAM, B.; CAVERLEY, N. Factors in absenteeism and presenteeism: life events and health events. **Management Research News**, v.31, n.8, p.607-615, 2008.

- MARTENS, D.; BAESENS, B.; FAWCETT, T. Editorial survey: swarm intelligence for data mining. **Machine Learning**. v.82, n.1, p.1-42, 2011.
- MARTINEZ, L. F. & FERREIRA, A. I. Sick at work: presenteeism among nurses in a Portuguese public hospital. **Stress e Health**, v.28, n.4, p.297-304, 2011.
- MARTINIANO, A.; FERREIRA, R. P.; SASSI, R. J.; AFFONSO, C. Application of a neuro fuzzy network in prediction of absenteeism at work. **In Information Systems and Technologies (CISTI)**, 7th Iberian Conference on, p.1-4. IEEE, june, 2012.
- MARTINIANO, A.; FERREIRA, R. P.; SASSI, R. J. **Mineração de dados aplicada a um estudo sobre ergonomia em uma empresa de logística: conhecimento e prática**. XV SEGeT – Simpósio de Excelência em Gestão e Tecnologia. Resende. Rio de Janeiro, p.1-16, 2018.
- MATTAR, F. N. **Pesquisa de marketing: metodologia, planejamento**. 7ª ed. Rio de Janeiro: Campus, 2014.
- MBP. Multiple Back-Propagation. Disponível em: <<http://mbp.sourceforge.net/>>. Acesso em: 29 out. 2019.
- MESQUITA, F. L. F. M.; MESQUITA, A. U. **Perfil antropométrico, socioeconômico e de saúde de funcionários assistidos pelo programa de alimentação do trabalhador**. Revista Brasileira em Promoção da Saúde, v.26, n.2, p.192-200, 2013.
- MEZIAT FILHO, N.; SILVA, G. A. **Invalidez por dor nas costas entre segurados da Previdência Social do Brasil**. Rev Saúde Pública, v.45, n.3, p.494-502, 2011.
- MIRAGLIA, M.; JOHNS, G. A Multi-Level Integration of the Antecedents and Dynamics of Absenteeism. In: Academy of Management Proceedings. Briarcliff Manor, NY 10510: **Academy of Management**, p.1-40, 2018.
- MITCHELL, T. M. Machine Learning. **McGraw-Hill**, 1997.
- NADADUR, G.; PARKINSON, M. B. The role of anthropometry in designing for sustainability. **Ergonomics**, v.56, n.3, p.422-439, 2013.
- NAIK, A.; SAMANT, L. Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, v.85, p. 662-668, 2016.
- NORTON, K. OLDS, T. **Antropométrica**. Editora Artmed. 1ª ed., 2005.
- OCHOA, A.; HERNANDEZ, A.; SANCHEZ, J.; GARCIA, Y.; DUMONT, M.; IZTEBEGOVIC, H. Identify an Adequate Anthropometry to Water Polo Using Social Data Mining. In: 2009 **International Conference on Electrical, Communications, and Computers**. IEEE, p.144-147, 2009.
- OLIVEIRA, M. M.; ANDRADE, S. S.C. A.; SOUZA, C. A. V.; PONTE, J. N.; SZWARCOWALD, C. L.; MALTA, D. C. M. **Problema crônico de coluna e diagnóstico de distúrbios osteomusculares relacionados ao trabalho (DORT) autorreferidos no Brasil: Pesquisa Nacional de Saúde, 2013**. Epidemiol. Serv. Saúde, 287 Brasília, v.24, n.2, p.287-296, 2015.
- OTERO, F. E. B. MYRA: a Java ant colony optimization framework for classification algorithms. In: **Proceedings of the Genetic and Evolutionary Computation Conference Companion**. ACM, p.1247-1254, 2017.

- OTERO, F. E. B.; FREITAS, A. A. Improving the interpretability of classification rules discovered by an ant colony algorithm. **In Proceedings of the 15th annual conference on Genetic and evolutionary computation**. ACM, p. 73-80, 2013.
- OTERO, F. E. B.; FREITAS, A. A.; JOHNSON, C. G. A new sequential covering strategy for inducing classification rules with ant colony algorithms. **IEEE Transactions on Evolutionary Computation**, v.17, n1, p.64-76, 2013.
- ÖZMEN, M.; AYDOĞAN, E. K.; DELICE, Y.; TOKSARI, M. D. Churn prediction in Turkey's telecommunications sector: A proposed multiobjective–cost-sensitive ant colony optimization. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, p.1-23, 2019.
- PAAKAALA, A. V. R.; MACHA, S. S.; MUDIGONDA, K. S. Evaluation of Clustering Algorithms on Absenteeism at Work Dataset. **IJSRD - International Journal for Scientific Research & Development**, v.6, Issue 06, p.337-342, 2018.
- PAN, S.; ZHOU, S. Evaluation Research of Credit Risk on P2P Lending based on Random Forest and Visual Graph Model. **Journal of Visual Communication and Image Representation**, p.1-8, 2019.
- PARPINELLI, R. S.; LOPES, H. S.; FREITAS, A. A. Data mining with an ant colony optimization algorithm. **IEEE Transactions on Evolutionary Computation**, v.6, n.4, p.321–332, 2002.
- PASCHOARELLI, L. C.; SANTOS, M. M. **Design e ergonomia: aspectos tecnológicos**. SciELO-Editora UNESP, 2009.
- PATEL, S.; PATEL, A. Performance Analysis and Evaluation of Clustering Algorithms. **International Journal of Innovative Technology and Exploring Engineering (IJITEE)**, v.8, p.179-183, 2019.
- PATTON, E.; JOHNS, G. Context and the social representation of absenteeism: Absence in the popular press and in academic research. **Human Relations**, v. 65, n.2, p. 217-240, 2012.
- PIATETSKY-SHAPIRO, G.; MATHEUS, C. J.; CHAN, P. K. Systems for Knowledge Discovery in Data bases. **IEEE Transactions on Knowledge and Data Engineering**, v.5, p.903-912, 1993.
- PICHLER, Stefan; ZIEBARTH, Nicolas R. The pros and cons of sick pay schemes: Testing for contagious presenteeism and noncontagious absenteeism behavior. **Journal of Public Economics**, v.156, p.14-33, 2017.
- PIMENTEL, J. S.; DIAS, J. L. F.; ALMEIDA, T. C. D.; FREITAS, A. R. **Pode a ginástica laboral influenciar sobre os determinantes musculoesqueléticos relacionado à prevenção de distúrbios osteomusculares relacionados ao trabalho?** Revista Ação Ergonômica, v. 12, n.1, p.115-122, 2017.
- PINTO, A. G.; TERESO, M. J. A.; ABRAHÃO, R. F. A. **Práticas ergonômicas em um grupo de indústrias da Região Metropolitana de Campinas: natureza, gestão e atores envolvidos**. Rev.Gest. Prod., São Carlos, v.25, n.2, p.398-409, 2018.
- PIZO, C. A.; MENEGON, N. L. **Análise ergonômica do trabalho e o reconhecimento científico do conhecimento gerado**. **Ergonomia da Atividade Compartilhamento de Material sobre Ergonomia e Trabalho Humano**. Produção, v.20, n.4, p.657-668, 2010.
- PRATER, T.; SMITH, K. Underlying factors contributing to presenteeism and absenteeism. **Journal of Business & Economics Research**, v.9, n.6, p.1-14, 2011.

PRIYANKA, D.; NAYAK, J. Empirical Analysis of Absenteeism at Work Place Using Machine Learning. In: **International Conference on Application of Robotics in Industry using Advanced Mechanisms**. Springer, Cham, p.150-160, 2019.

QOMARIYAH, N.; SUCAHYO, Y. G. Employees attendance patterns prediction using classification algorithm case study: a private company in Indonesia. **Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE)**, v.1, p.2349-1469, 2014.

QUICK, T.C.; LAPERTOSA, J. **Análise do absentismo em usina siderúrgica**. Revista brasileira de saúde ocupacional, v.10, n.40, p.62-7, 1982.

RABBI, B.; RABBI, D. B. K.; GONÇALVES, V. S.; JÚNIOR, E. R. G.; BRASIL, J. A. **Mineração de dados aplicada a base de reclamações sobre produtos e serviços do programa de proteção e defesa do consumidor/Data mining applied on products and services claims of the consumer protection and defense program**. Brazilian Journal of Development, v.4, n.5, p.1689-1701, 2018.

RAJAB, S.; SHARMA, V. A review on the applications of neuro-fuzzy systems in business. **Artificial Intelligence Review**, v. 49, n.4, p.481-510, 2018.

RAMSAY, C. M.; OGULEDO, V. I.; KRUTTO, A. Exploring the Optimal Design of an Employer-Sponsored Sickness-Disability Compensation Insurance Plan When Sickness Presenteeism Is Penalized. **North American Actuarial Journal**, v.22, n.3, p.426-457, 2018.

ROTHSTEIN, J. R.; BERNDT, A.; SOUZA M.; J. C.; LANFERDINI, F. J. **Impacto de uma metodologia interativa de ergonomia de conscientização**. Fisioterapia e Pesquisa, v.20, n.1, p.11-16, 2013.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, v.323, n.6088, p.533-536, 1986.

SABRI, S. N.; SAIAN, R. Predicting flood in perlis using ant colony optimization. In **Journal of Physics: Conference Series**. IOP Publishing. v.855, n.1, p.1-8, 2017.

SADATH, L. Data Mining: A Tool for Knowledge Management in Human Resource. **International Journal of Innovative Technology and Exploring Engineering**, v.2, n.6, p.154-159, 2013.

SAMEH, A.; MAGDY, K. Data Mining Ant Colony for Classifiers. **International Journal of Basic & Applied Sciences**, v.10, n.3, p.28-35, 2010.

SANTOS, R., W.; BRRATZ, D.; TONIN, L. A.; MENEGON, N. L. **Análise do uso integrado de um sistema de captura de movimentos com um software de modelagem e simulação humana para incorporação da perspectiva da atividade**. Rev. Gest. Prod., São Carlos, v.23, n.3, p.612-624, 2016.

SCHLOSSER, J. F.; SCHLOSSER, J. F.; DEBIASI, H.; PARCIANELLO, G.; RAMBO, L. **Antropometria aplicada aos operadores de tratores agrícolas**. Ciência Rural, v.32, n.6, p.983-988, 2002.

SECRETARIA DE PREVIDÊNCIA, Ministério da Fazenda. Saúde do trabalhador: Dor nas costas foi doença que mais afastou trabalhadores em 2017. Disponível em: < <http://www.previdencia.gov.br/2018/03/saude-do-trabalhador-dor-nas-costas-foi-doenca-que-mais-afastou-trabalhadores-em-2017/> >. Acessado 19 jan. 2020.

SIEGEL S, CASTELLAN N. Nonparametric Statistics for the Behavioral Sciences, 2ª.ed. New York: **McGraw-Hill**, p 284-285, 1988.

- 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.
- SHIMABUKU, R. H.; MENDONÇA, H.; FIDELIS, A. (2017). **Presenteísmo: contribuições do modelo demanda-controle para a compreensão do fenômeno.** *Cad. psicol. soc. trab.*, v.20, n.1, p.65-78, 2017.
- SHUANG, Y.; JIANGJIANG, X.; ZHONGWEI, Y.; ANZHI, Z.; XIA, Y.; GUAN, D.; YAKUN, L. Loss of work productivity in a warming world: differences between developed and developing countries. **Journal of Cleaner Production**, v.208, p.1219-1225, 2019.
- SILVA, A. C. C. J.; FURTADO, J. H.; ZANINI, R. R. Um estudo sobre a qualidade de vida no trabalho (QVT) e os fatores associados. **Iberoamerican Journal of Industrial Engineering**, v.7, n.14, p.182-200, 2015.
- SILVA, D. M. P. P.; MARZIALE, M. H. P. Problemas de saúde responsáveis pelo absenteísmo de trabalhadores de enfermagem de um hospital universitário. *Acta Scientiarum. Health Sciences*, v.25, n.2, p.191-197, 2003.
- SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à Mineração de dados: com aplicações em R.** Elsevier. 1ª. ed. Rio de Janeiro, 2016.
- SILVA, I. N.; SPATTI, D.H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e Ciências Aplicadas.** Artliber Editora. 2ª ed. São Paulo, 2016.
- SILVA, M. M.; CAVEIÃO, C. **Análise dos afastamentos de saúde dos trabalhadores de ensino de Divinópolis–MG.** *Revista Saúde e Desenvolvimento*, v.10, n.5, p.138-156, 2016.
- SILVEIRA, M. A.; Kikuchi, L. S.; Lima, A. S.; Silveira, R. D. **Inovação e aprendizagem organizacional: Abordagem TCD para desenvolvimento de competências em empresa do setor eletrônico.** In Congresso Latino-Iberoamericano de Gestão de Tecnologia, v.16, p.1-19, 2015.
- SIMÕES, M. G.; SHAW, I. S. *Controle e Modelagem fuzzy.* São Paulo: Blucher: FAPESP, 2ª ed., 2007.
- SOUZA, R. L. R.; CAMAROTTO, J. A. **Tecnologias na coleta seletiva no Brasil: uma perspectiva antropotecnológica.** *Revista Ação Ergonômica*, v.12, n.1, p.42-47, 2017.
- SRIVASTAVA, R.; TIWARI, A. N.; GIRI, V. K. Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India. **Heliyon**, v.5, n. 10, p.1-14, 2019.
- SRUAMSIRI, R.; MORI, Y.; MAHLICH, J. Productivity loss of caregivers of schizophrenia patients: a cross-sectional survey in Japan. **Journal of Mental Health**, v.27, n.6, p.583-587, 2018.
- SUTALAKSANA, I. Z.; WIDYANTI, A. *Anthropometry approach in workplace redesign in Indonesian Sundanese roof tile industries.* **International Journal of Industrial Ergonomics**, v.53, p.299-305, 2016.
- SUZUKI, T.; MIYAKI, K.; SONG, Y.; TSUTSUMI, A.; KAWAKAMI, N.; SHIMAZU, A.; KURIOKA, S. Relationship between sickness presenteeism (WHO–HPQ) with depression and sickness absence due to mental disease in a cohort of Japanese workers. **Journal of affective disorders**, v.180, p.14-20, 2015.

- TEWARI, K.; VANDITA, S.; JAIN, S. Predictive Analysis of Absenteeism in MNCS Using Machine Learning Algorithm. In: **Proceedings of ICRIC 2019**. Springer, Cham, 2020. p. 3-14, 2020.
- TROUNSON, J.; PFEIFER, J. Corrections officer wellbeing: Training challenges and opportunities. **Practice**, v.5, n. 1, p.22-28, 2017.
- TROUNSON, J. S.; PFEIFER, J. E.; SKUES, J. L. Perceived workplace adversity and correctional officer psychological well-being: an international examination of the impact of officer response styles. **The Journal of Forensic Psychiatry & Psychology**, v. 30, n. 1, p. 17-37, 2018.
- UMANN, J.; GUIDO, L. de A.; SILVA, R.M. Estresse, *coping*, e presentismo em enfermeiros que assistem pacientes críticos e potencialmente críticos. **Rev Esc Enferm USP**, São Paulo, v.48, n.5, p.891- 898, 2014.
- UTHAYAKUMAR, J.; METAWA, N.; SHANKAR, K.; LAKSHMANAPRABU, S. K. (2020). Financial crisis prediction model using ant colony optimization. **International Journal of Information Management**, v.50, p.538-556, 2020.
- VESANTO, J.; ALHONIEMI, E. Clustering of the Self-Organizing Map. **IEEE Transactions on Neural Networks**, v.11, n.2, p.586-600, 2000.
- VISCOVERY. Viscovery® SOMine. Disponível em: <<https://www.viscovery.net/>>. Acesso em: 15 jan. 2020.
- WAHID, Z.; SATTER, A. K. M.; AL IMRAN, A.; BHUIYAN, T. Predicting Absenteeism at Work Using Tree-Based Learners. In **Proceedings of the 3rd International Conference on Machine Learning and Soft Computing**, p.7-11, 2019.
- WAN, H. C.; DOWNEY, L. A.; STOUGH, C. Understanding non-work presenteeism: Relationships between emotional intelligence, boredom, procrastination and job stress. **Personality and Individual Differences**, v.65, p.86-90, 2014.
- WEKA. *Waikato Environment for Knowledge Analysis*. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 19 jan. 2020.
- WISNER, A. A antropotecnologia. **Estudos Avançados**, São Paulo, v.6, n.16, p.29-34, 1992.
- WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. Data Mining: Practical machine learning tools and techniques. Cambridge. **Morgan Kaufmann**, 2016.
- WHYSALL, Z.; BOWDEN, J.; HEWITT, M. Sickness presenteeism: measurement and management challenges. **Ergonomics**, v.6, n.3, p.341-354, 2018.
- XIONG, Z.; CUI, Y.; LIU, Z.; ZHAO, Y.; HU, M.; HU, J. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. **Computational Materials Science**, v.171, p.1-12, 2020.
- YIN, R. K. Estudo de Caso - Planejamento e Métodos. **Bookman Editora**. 5ª ed. Porto Alegre, 2015.
- YOSHIKAWA, M.; OTANI, K. Ant Colony Optimization Routing Algorithm with Tabu Search. **Proceedings of the Internacional MultiConference of Engineers and Computer Scientists**, 2010. V. III. IMECS 2010. Hong Kong, 2010.
- ZAMANIAN, Z.; DANESHMANDI, H.; SETOODEH, H.; NAZARIPOOR, E.; HAGHAYEGH, A.; SARVESTANI, S. S. Risk Assessment of Musculoskeletal Disorders and

Determination of the Associated Factors among Workers of a Dairy Products Factory. **Journal of Health Sciences and Surveillance System**, v.2, n.4, p.134-139, 2014.

ZAVARIZZI, C. P.; ALENCAR, M. C. Baracho. **Afastamento do trabalho e os percursos terapêuticos de trabalhadores acometidos por LER/Dort**. Saúde em Debate, v.42, p.113-124, 2018.

ZHOU, Z. Q.; MARTINEZ, L. F.; FERREIRA, A. I.; RODRIGUES, P. Supervisor support, role ambiguity and productivity associated with presenteeism: A longitudinal study. **Journal of Business Research**, v.69, n.9, p.3380-3387, 2016.

Anexo A – Certificado de Registro de Programa de Computador



REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512018001297-3**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 14/01/2018, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: Software Tool for Absenteeism Prediction

Data de criação: 14/01/2018

Titular(es): ASSOCIAÇÃO EDUCACIONAL NOVE DE JULHO

Autor(es): RENATO JOSÉ SASSI; RICARDO PINTO FERREIRA; ANDRÉA MARTINIANO DA SILVA

Linguagem: C

Campo de aplicação: AD-02

Tipo de programa: IA-02

Algoritmo hash: SHA-512

Resumo digital hash:

0ad5657fafa7b1cdd5b51239c7a345d824d7e408b351d41833dfda2efd691560f40d550aec07781c47419565da1ecfba64da58dfbfa507c9f86c165f067e0c5a

Expedido em: 31/07/2018

15 de Novembro
REPÚBLICA FEDERATIVA DO BRASIL
de 1889