

**FACULDADE DE TECNOLOGIA DE SÃO PAULO**

**Gabriela Beatriz Amadori**

**METODOLOGIA DE PESQUISA DE MINERAÇÃO DE DADOS VOLTADA À  
MELHORA DA ERGONOMIA QUE AS EMPRESAS OFERECEM AOS  
FUNCIONÁRIOS**

**SÃO PAULO**

**2022**

**FACULDADE DE TECNOLOGIA DE SÃO PAULO**

**Gabriela Beatriz Amadori**

Trabalho submetido como exigência parcial  
para a obtenção do Grau de Tecnólogo em  
Análise e Desenvolvimento de Sistemas  
Orientador: Me. Edson Roberto Barbosa Ceroni

**SÃO PAULO**

**2022**

**FACULDADE DE TECNOLOGIA DE SÃO PAULO**

**Gabriela Beatriz Amadori**

**METODOLOGIA DE PESQUISA DE MINERAÇÃO DE DADOS VOLTADA À  
MELHORA DA ERGONOMIA QUE AS EMPRESAS OFERECEM AOS  
FUNCIONÁRIOS**

Trabalho submetido como exigência parcial para a obtenção do Grau de  
Tecnólogo em Análise e Desenvolvimento de Sistemas.

Parecer do Professor Orientador

---

---

---

Conceito/Nota Final: \_\_\_\_\_

**Atesto o conteúdo contido na postagem do ambiente TEAMS pelo aluno e  
assinada por mim para avaliação do TCC.**

Orientador: Me. Edson Roberto Barbosa Ceroni

SÃO PAULO, 22 de junho de 2022.

Assinatura do Orientador

Assinatura do aluno

Ao meu irmão Flavio Q.E.P.D.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus por ter me dado a oportunidade de fazer minha terceira graduação, agradeço aos meu esposo por ter me ajudado e ter me brindado todo apoio e paciência para chegar até aqui, agradeço ao meu irmão Cesar por ter me impulsado a realizar este curso, agradeço aos meus pais, minha filha, neta, a todos que contribuíram diretamente ou indiretamente e ao meu orientador Me. Edson Roberto Barbosa Ceroni por ter me dado todo o suporte para realizar meu trabalho.

"Nenhuma grande descoberta foi feita jamais sem um palpite ousado."

Isaac Newton

## RESUMO

De acordo com a Organização Mundial da Saúde (OMS), em torno de 80% da população está sujeita às famosas “dores nas costas”. A dor lombar é a segunda causa principal que leva as pessoas às consultas médicas. (“Quais as profissões mais sujeitas a problemas na coluna?”) Em geral, alguns profissionais estão mais propensos do que outros a sofrerem com problemas na coluna. Entre esses profissionais estão os que trabalham com tecnologia, dado que ficam muitas horas sentados frente ao computador. Este trabalho traz a discussão sobre a possibilidade de utilização das técnicas de Data Mining através da coleta de dados por parte das empresas, para proporcionar uma melhor qualidade na saúde dos seus colaboradores, em relação aos distúrbios osteomusculares. Para auxiliar nesta tarefa foram abordados os conceitos fundamentais que poderão ser aprofundados e aplicados pelo usuário para desenvolver um modelo de previsão que melhor se ajuste aos dados coletados e assim responder ao problema apresentado. Como técnica de pesquisa foi implementado um questionário entre os profissionais da área em questão. De acordo com a análise, a escassez dos dados não permitiu chegar a uma conclusão decisiva para comprovação da discussão presente. Acredita-se que a obtenção de um maior volume de dados possa contribuir para a alcançar resultados relevantes.

**Palavras-Chave:** Dados, Distúrbios osteomusculares, Ergonomia, Tecnologia

## ABSTRACT

According to the World Health Organization (WHO), it is a fact that about 80% of the world population suffer from back pain. The second main cause that leads people to seek medical advice is the lower back pain. (¿“Which professions are most prone to back problems?”) In general, some professionals are more likely than others to suffer column issues. Among them are especially those who work in technology who, in fact, spend many hours sitting in front of their computer. This paper opens the debate about the possibility of usage of data mining techniques through the data collection from companies to provide, for example, a better life quality to their employees regarding the musculoskeletal disorders. In order to help with this task, main concepts were addressed. They might be studied in more detail and applied by the user to develop a forecast model based on the data collection in which case will answer the present problem. In effect, a survey was implemented as the data collection method completed by technology professionals. Regarding the data analysis, the lack of sufficient data prevented from getting a relevant conclusion to confirm the present debate. Actually, it is known that a big data gathering would contribute to the process of getting relevant results.

**Keywords:** Data, Musculoskeletal disorders, Ergonomics, Technology



## LISTA DE FIGURAS

Figura 1 - KDD .....	16
Figura 2 - CRISP-DM .....	18
Figura 3 - MACHINE LEARNING .....	19
Figura 4 - APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO .....	20
Figura 5 - MODELO DE CLASSIFICAÇÃO .....	21
Figura 6 - MATRIZ DE CONFUSÃO .....	22
Figura 7 - UNDERFITING .....	23
Figura 8 - OVERFITING .....	23
Figura 9 - REGRESSÃO LINEAR .....	24
Figura 10 - REGRESSÃO LOGÍSTICA .....	25
Figura 11 - ÁRVORE DE DECISÃO .....	26
Figura 12 - MÁQUINAS DE VECTOR DE SUPORTE .....	26
Figura 13 - CLASSIFICADOR NAYVE BAYES .....	27
Figura 14 - VIZINHO MAIS PRÓXIMO .....	28
Figura 15 - FLORESTAS ALEATÓRIAS .....	28
Figura 16 - TÉCNICA DE AGRUPAMENTO .....	29
Figura 17 - TÉCNICA DE ASSOCIAÇÃO .....	30
Figura 18 - SUPORT CONFIANÇA LIFT .....	31
Figura 19 - ALGORITMO APRIORI .....	32
Figura 20 - APRENDIZADO POR REFORÇO .....	32
Figura 21 - MAPA DE RESPOSTAS .....	36
Figura 22 - ATIVIDADE POR PAÍS .....	36
Figura 23 - ATIVIDADE ATUAL .....	37
Figura 24 - SEXO .....	37
Figura 25 - FREQUÊNCIA DE IDADE .....	38
Figura 26 - ATIVIDADE X DOR .....	39
Figura 27 - ATIVIDADE FÍSICA X ATIVIDADE ATUAL .....	39
Figura 28 - HORA/DIA/COMPUTADOR .....	40
Figura 29 - DOR NAS COSTAS .....	40
Figura 30 - DOR ANTES .....	41
Figura 31 - DOR DEPOIS .....	41
Figura 32 - ALTURA TELA X DOR .....	42

Figura 33 - HORAS/DIA X DOR .....	43
Figura 34 - HÁ QUANTOS ANOS TRABALHA? .....	43
Figura 35 - TRABALHO X DOR .....	44
Figura 36 - SOLICITAR EXAMES .....	45
Figura 37 - FAZ GINÁSTICA LABORAL .....	45

## LISTA DE TABELAS

Tabela 1 - ATIVIDADE ATUAL.....	37
Tabela 2 - SEXO .....	37
Tabela 3 - FREQUÊNCIA DE IDADE .....	38
Tabela 4 - ATIVIDADE X DOR .....	38
Tabela 5 - ATIVIDADE FÍSICA X ATIVIDADE ATUAL .....	39
Tabela 6 - HORA/DIA/COMPUTADOR .....	40
Tabela 7 - DOR NAS COSTAS.....	41
Tabela 8 - DOR ANTES.....	41
Tabela 9 - DOR DEPOIS.....	41
Tabela 10 - ALTURA TELA X DOR.....	42
Tabela 11 - HORAS/DIA X DOR .....	42
Tabela 12 - HÁ QUANTOS ANOS TRABALHA?.....	43
Tabela 13 – TRABALHO X DOR.....	44
Tabela 14 - SOLICITAR EXAMES.....	44
Tabela 15 - FAZ GINÁSTICA LABORAL.....	45

## SUMÁRIO

1. INTRODUÇÃO .....	13
2. A MINERAÇÃO DE DADOS – Data Mining – DM .....	15
2.1 O KDD .....	15
2.2 O CRISP-DM .....	17
3. O APRENDIZADO DE MÁQUINA OU MACHINE LEARNING – ML .....	19
3.1 Aprendizado Supervisionado .....	20
3.1.1 Classificação e Regressão.....	20
3.1.2 Regressão Linear.....	24
3.1.3 Regressão Logística .....	25
3.1.4 Árvore de Decisão.....	25
3.1.5 Máquinas de vetor de suporte.....	26
3.1.6 Classificador Naïve Bayes .....	27
3.1.7 Vizinho mais próximo .....	27
3.1.8 Florestas aleatórias.....	28
3.2 Aprendizado Não Supervisionado.....	29
3.2.1 Técnica de Agrupamento ou Clusterização .....	29
3.2.2 Técnica de Associação .....	30
3.3 Aprendizado por reforço ou Reinforcement Learning .....	32
5. METODOLOGIA DA PESQUISA .....	34
6. RESULTADOS DA PESQUISA.....	36
7. DISCUSSÃO SOBRE OS RESULTADOS DA PESQUISA .....	47
8. CONSIDERAÇÕES FINAIS .....	49
REFERÊNCIAS.....	50

## 1. INTRODUÇÃO

Esta pesquisa trata sobre a possibilidade da utilização das técnicas de mineração de dados aplicada a um grande volume de informações coletadas pelas empresas, para poder proporcionar uma melhoria na qualidade de trabalho dos colaboradores do segmento de tecnologia. Esta melhoria está intimamente relacionada aos problemas osteomusculares.

O objetivo desse trabalho é apresentar os conceitos básicos das técnicas de Data Mining e encontrar alguma correspondência entre os problemas que os colaboradores apresentam, com a respectiva função exercida, através dos dados coletados no questionário.

Diante das informações oferecidas pela Organização Mundial da Saúde, relacionadas à alta incidência de consultas médicas por “dores nas costas” e que os trabalhadores do segmento de tecnologia têm mais propensão a este problema devido ao tempo que passam sentados frente ao computador, levantaram-se as seguintes questões a serem analisadas no decorrer deste trabalho: Os equipamentos oferecidos pelas empresas para desenvolver trabalhos nesta área, são ergonômicos? É oferecida a prática de ginástica laboral para evitar os períodos longos frente ao computador? São realizados exames prévios para avaliar as condições osteomusculares dos colaboradores? As empresas realizam avaliações periódicas para detecção de possíveis problemas gerados pelo exercício da atividade desenvolvida? Existem incentivos à prática de atividade física para assim evitar o sedentarismo que o exercício da profissão implica? Qual é o tamanho da preocupação por parte das empresas sobre a problemática apresentada?

Para tais questionamentos a hipótese provável é a falta de informação suficiente que possa comprovar a correlação da atividade desenvolvida pelos colaboradores, com os problemas de saúde apresentados.

Neste trabalho buscou-se aplicar uma pesquisa qualitativa e quantitativa ressaltando os aspectos qualitativos ao que se refere as condições de execução das atividades relacionadas ao segmento em questão. Também se fez uso da pesquisa quantitativa, a partir de questionários fechados e anônimos, para analisar as condições das “dores nas costas” da população que estuda ou trabalha com tecnologia e as condições em que são realizadas as atividades. Para tal, aplicou-se

384 questionários em diferentes pontos dos países pertencentes ao continente americano, com um maior número coletado no Brasil e, adiante, tabularam-se os dados para compreender tais informações.

A motivação para analisar e compreender a relação entre os problemas de saúde osteomusculares da população que exerce atividade relacionada com tecnologia e a condição da execução da atividade em si, se deu a partir da urgência, desta pesquisadora, em entender os próprios sintomas decorrentes da atividade em questão. Aplica-se esta preocupação à toda população que desenvolve esta atividade, tornando-se de suma importância pesquisar e refletir sobre esta problemática dado o aumento exponencial da tecnologia e a necessidade de profissionais para ocupar estas novas vagas de emprego.

No primeiro capítulo serão abordados os conceitos fundamentais da mineração de dados. No segundo capítulo serão apresentadas as diferentes técnicas de mineração de dados. No terceiro capítulo, destaca-se a metodologia utilizada na pesquisa e a discussão dos resultados obtidos sobre os dados coletados. No quarto e último capítulo a conclusão.

## **2. A MINERAÇÃO DE DADOS – DATA MINING – DM**

Atualmente as tecnologias estão se voltando para a coleta, armazenamento e processamento de grandes volumes de dados, chamados de Big Data. Esses dados precisam ser analisados para que, a partir deles, possa se extrair informações e conhecimentos úteis para a tomada de decisão frente as diversas necessidades.

As técnicas para análise de Big Data permitem que as empresas consigam atingir seus objetivos nos negócios e isto é fundamental para obter um crescimento e expansão bem-sucedidos (IJDSR, 2019)

Dentro da estrutura de Big Data temos os chamados 4 Vs: Velocidade, Variedade, Volume e Veracidade, portanto, o Big Data permite lidar com grande volume de dados a grande velocidade, trazendo uma análise de confiança mesmo com dados bem diversos, proporcionando um importante auxílio na mineração de dados.

### **2.1 O KDD**

A mineração de dados é um conjunto de técnicas e algoritmos utilizados em alguns processos da estrutura da análise de dados, como o KDD (Knowledge Discovery in Databases - Descoberta de conhecimentos a partir de base de dados). A partir da mineração de dados é possível explorar dados em busca de padrões consistentes, para detectar relacionamentos sistemáticos entre variáveis e assim detectar novos subconjuntos de dados.

O KDD é um processo composto por várias etapas, para se atingir um conhecimento a partir de uma base de dados. A diferença entre ele e o fluxo comum [dado - informação - conhecimento] é que no KDD o dado não precisa ser transformado em informação, o conhecimento é extraído diretamente da base de dados com modelos, análises estatísticas e interpretações de relações entre os dados. Esse processo de criação de conhecimento a partir de dados ocorre pela criação de diferentes modelos, frutos da mineração de dados.

Apesar de parecer tentadora a ideia de se “extrair vastos conhecimentos de uma base de dados” esse processo é bem custoso, exige uma boa base de dados, métodos de validação dos modelos e profissionais capacitados para entenderem tais modelos, relatórios e dashboards para somente então tomar uma decisão.

As etapas do KDD (figura 1) são:

- Seleção: etapa responsável pela seleção de quais dados do banco de dados devem ou não entrar. Só serão usados os dados que fizerem sentido para o modelo a ser criado.

- Pré-Processamento: etapa responsável pela limpeza dos dados, onde serão removidos dados nulos, dados inconsistentes ou anomalias. Dados ruins distorcem o modelo e podem gerar uma interpretação equivocada.

- Transformação: etapa responsável por transformar os dados limpos para reduzir o escopo, facilitar o processamento, reduzir custos e trazer melhores resultados. Dessa forma, os dados serão normalizados, agregados ou reduzidos para melhor atender à criação do modelo.

- Mineração de Dados: etapa na qual serão aplicadas técnicas e algoritmos sobre os dados para a efetiva criação do modelo. Uma vez criado este, especialistas realizarão uma avaliação para validá-lo, confirmando se ele é capaz de responder uma hipótese, seja ela verdadeira ou falsa.

- Interpretação e avaliação: etapa final que envolve a própria avaliação do modelo produzido. Esta validação pode ser realizada por meios estatísticos ou por meio de profissionais que tenham conhecimento do caso de uso para poder confirmar que o modelo trará respostas consistentes. E detentora do modelo poderá aplicá-lo diversas vezes, sempre que precisar confirmar uma hipótese em uma nova base de dados.

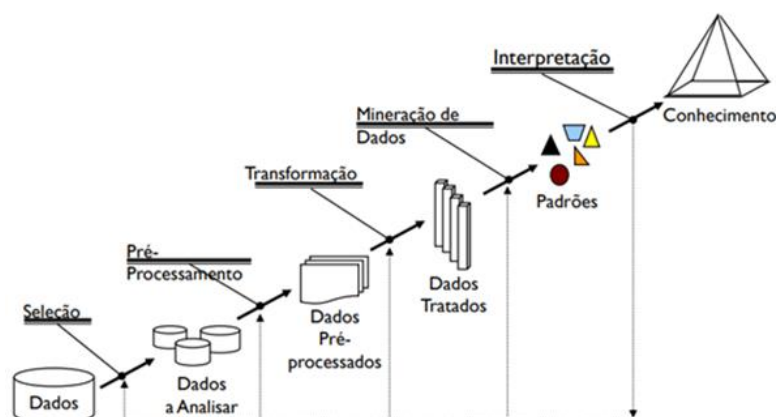


Figura 1 - KDD



## 2.2 O CRISP-DM

O CRISP-DM ou processo padrão entre indústrias de mineração de dados é um modelo de processo para padronizar as fases de implementação de mineração de dados. É um processo iterativo, ou seja, deve ocorrer N vezes até que se encontre o modelo ideal para atender a uma demanda ou necessidade.

As etapas do CRISP-DM são:

- **Compreensão do negócio:** é um processo gerencial e menos técnico que aborda o estudo do contexto, demandas e viabilidades do projeto. Por sua vez, se divide em quatro atividades: determinação dos objetivos do projeto, análise de viabilidade, objetivos do Data Mining e planejamento de implementação do Data Mining.

- **Compreensão dos dados:** trata-se de identificar, coletar e analisar os grupos de dados que serão utilizados para atingir o objetivo do projeto. Também se divide em quatro processos: coletar dados iniciais, descrever os dados, explorar possibilidades e relacionamentos entre os dados e verificar a qualidade dos dados.

- **Preparação dos dados:** etapa na que ocorre o processamento dos dados que serão necessários para o sucesso da etapa seguinte. Se divide em cinco atividades: seleção de dados, limpeza de dados, construção dos dados, integração dos dados e formatação dos dados.

- **Modelagem:** nesta etapa são construídos modelos, usando os dados tratados anteriormente, aplicando técnicas de Data Mining que veremos posteriormente. Os processos realizados aqui são: seleção da técnica de modelagem, design de testagem, construção do modelo e validação do modelo. Este processo é iterativo até encontrar o modelo ideal.

- **Avaliação:** esta validação procura alinhar o modelo produzido e o modelo do projeto. Nesta etapa ocorrem três processos: a avaliação dos resultados, a revisão dos processos e a definição dos próximos passos.

- **Implementação ou distribuição:** etapa na que ocorre a verdadeira aplicação do modelo para atender as necessidades do projeto. É composta por quatro

processos: o planejamento do deploy (implantação), monitoramento e manutenção, produção de relatório final e revisão do projeto.

A sequência das fases não é rigorosa. Geralmente é necessário mover se para frente e para trás entre as diferentes fases como podemos observar na figura 2

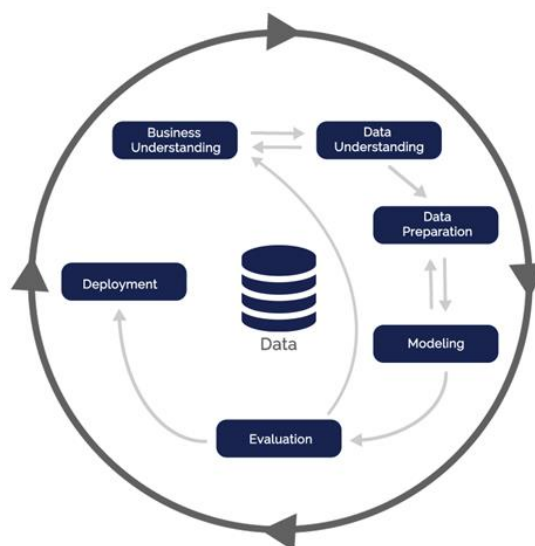


Figura 2 - CRISP-DM

As técnicas usadas pela mineração de dados compreendem basicamente a coleta, extração, análise e métodos estatísticos. É importante destacar que podem gerar descobertas de registros pouco comuns ou conexões não detectadas como também, desconhecidos padrões que podem resultar interessantes (IJDSR, 2019).

Para uma boa implementação de Data Mining contamos com três áreas fundamentais de estudo: a estatística, ciência preocupada em medir, a matemática, ciência preocupada em calcular e a Inteligência Artificial por meio do Machine Learning ou Aprendizado de Máquina, provisionam os algoritmos corretos para categorização e classificação dos dados.

### 3. O APRENDIZADO DE MÁQUINA OU MACHINE LEARNING – ML

Historicamente o Aprendizado de Máquina surge na década de 60. Originalmente, com o objetivo de aprender padrões baseados nos dados, era basicamente de caráter computacional. Já no final da década de 90, ampliou-se para ser considerada um campo por si mesmo, começando a ter muita intervenção da estatística (IZBICKI e SANTOS, 2020).

O Machine Learning (ML) é a capacidade que as máquinas têm para aprender automaticamente sem serem explicitamente programadas, ou seja, o algoritmo tem a capacidade de adquirir conhecimento a partir de observações, aprender dos dados, para melhorar, descrever e predecir resultados. Não existe um algoritmo único que funcione melhor para todos os problemas a resolver.

A mineração de dados é essencial para alimentar ferramentas de Machine Learning uma vez que proporcionam dados para treinar a máquina.

É importante ressaltar que apesar da mineração de dados produzir modelos, a decisão deve ser tomada pelas pessoas que estudam tais modelos, além disso existem limitações com base na estrutura de dados que está sendo abordada.

Estruturas complexas como Objetos, Fluxos de Dados ou Relacionamentos são questões difíceis de serem mineradas, seja por questão de processamento ou questões estruturais.

Dentro do ML temos diferentes tipos de aprendizados (figura 3): o Aprendizado Supervisionado onde apresentamos ao algoritmo dados de entrada já rotulados e suas respectivas saídas, o Aprendizado Não Supervisionado onde apresentamos somente os dados de entrada e o algoritmo descobre a saída e por último o Aprendizado por Reforço onde o componente aprende a reagir a uma recompensa (usado em jogos e robótica).

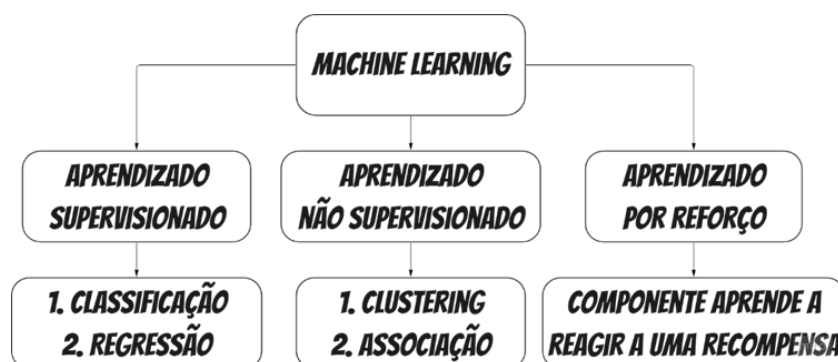


Figura 3 - MACHINE LEARNING

### 3.1 Aprendizado Supervisionado

O Aprendizado Supervisionado tem como objetivo fazer previsões aprendendo sobre a observação de um conjunto de dados rotulados (IZBICKI e SANTOS, 2020).

Os dados rotulados e conhecidos por meio de um algoritmo preditivo, consegue construir um modelo para estimar, sobre novos dados, o rótulo. Esse rótulo é o atributo de saída ou classe do dado. Desta forma a saída é conhecida (BERTOZZO, 2019).

Temos, portanto, uma parte dos dados utilizada para treinamento que irão gerar o modelo e uma parte dos dados utilizada para teste onde o modelo, gerado pelo aprendizado dos dados de treinamento, será aplicado.



Figura 4 - APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO

#### 3.1.1 Classificação e Regressão

Na classificação são determinadas as classes dos objetos para prever um novo dado automaticamente. Por exemplo, pode-se classificar um cliente que solicita um crédito através de dados históricos de transações anteriores, determinando dessa forma a categoria.

A partir de uma base de dados (dados históricos ou Dataset) se selecionam 70% a 80% para treino. Com essa informação cria-se um Modelo que será utilizado para testar o 30% ou 20% restante dos dados e dessa forma pode se ter uma previsão. Essa previsão terá uma avaliação de desempenho. A avaliação permitirá ver se o modelo funciona corretamente. Se assim for, será utilizado para apresentar novos dados a esse modelo que gerará uma previsão.

A figura a seguir representa um modelo de classificação.

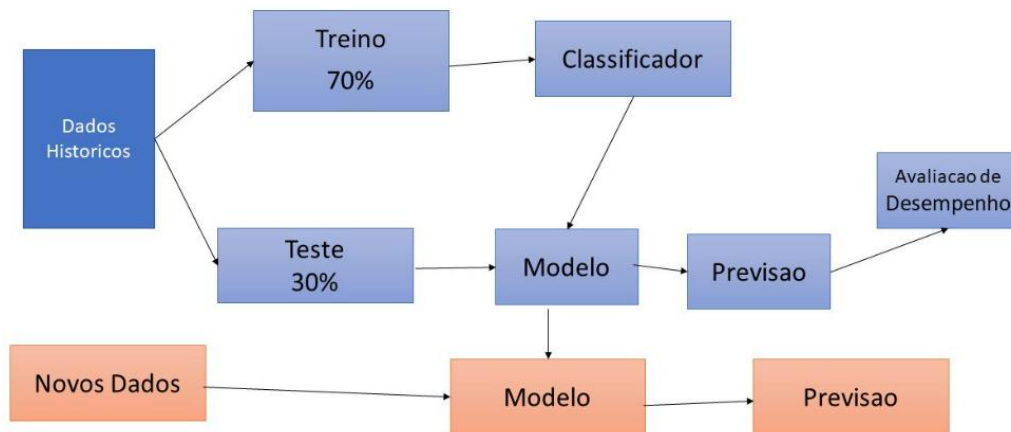


Figura 5 - MODELO DE CLASSIFICAÇÃO

A regressão também usa dados anteriores que já foram observados, porém a resposta é um valor numérico. Um exemplo de regressão pode ser considerado a previsão de um salário, usando um modelo que utilize a idade e os anos de escolaridade para fazer a análise. Na literatura encontram-se divergências no uso dos diferentes tipos de aprendizado supervisionado tanto para classificação como para regressão.

Como medida básica de desempenho utiliza-se a Acurácia. Quanto maior é a acurácia do modelo, maior serão os acertos e menores os erros cometidos. Para realizar o cálculo é utilizada a seguinte fórmula:

$$\text{Acurácia} = \frac{\text{Observações Classificadas Corretamente}}{\text{Número Total De Observações Classificadas}}$$

$$\text{Erro} = 1 - \text{Acurácia}$$

A acurácia também é aplicada a uma Matriz de Confusão (Confusion Matrix), na qual se exhibe, numa tabela, uma distribuição das classes atuais e suas previsões. Ao aplicar-se a fórmula da acurácia pode-se evidenciar a qualidade do modelo atual. A imagem a seguir representa uma Matriz de Confusão.

		Predição	
		P	N
Real	p	TP	FN
	n	FP	TN

Figura 6 - MATRIZ DE CONFUSÃO

- True Positive (TP): Valores classificados como positivos nos dados originais e corretamente previstos como positivos no modelo.
- False Positives (FP): Valores classificados como negativos nos dados originais e erroneamente previstos como positivos no modelo.
- "False Negatives (FN): Valores classificados como positivos nos dados originais e erroneamente previstos como negativos no modelo." ("Uma visão geral sobre machine learning – Classificação")
- True Negatives (TN): Valores classificados como negativos nos dados originais e corretamente previstos como negativos no modelo.

$$\text{Acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Existem dois conceitos muito importantes que devem ser abordados: o Underfitting (Sub-ajustado) e o Overfitting (Sobre-ajustado).

O Underfitting ocorre quando o modelo não conseguiu aprender o suficiente sobre os dados e leva a um erro elevado tanto nos dados de treino quanto nos dados de teste. A modelo falha por falta de amostras suficientes, portanto, não

generaliza o conhecimento. É incapaz de fazer uma relação entre variáveis de entrada e saída com precisão (IBM, 2021) A figura 6 (Platzi, 2022) representa um exemplo de Underfitting.



Figura 7 - UNDERFITTING

O Overfitting ocorre quando o modelo aprende demais sobre os dados, ou seja, o modelo é adequado apenas para os dados de treino, (apenas decora os dados de treino e não é capaz de generalizar para outros dados, nunca vistos antes). Quando isso acontece, os dados de treino apresentam resultados excelentes, enquanto a performance do modelo cai drasticamente com os dados de teste.

Pode-se observar na figura 7 (Platzi, 2022) que o modelo estatístico se ajusta exatamente a os dados de treinamento e por esse motivo o algoritmo não pode funcionar, ou seja, ele não serve para analisar os dados não vistos (IBM, 2021)



Figura 8 - OVERFITTING

Os diferentes tipos de Aprendizado Supervisionado são:

- Regressão Linear (Linear Regression)
- Regressão Logística (Logistic Regression)
- Árvores de decisão (Decision Trees)
- Máquinas de vetor de suporte (Support Vector Machines)
- Classificador Naïve Bayes (Naïve Bayes Classifier)
- Vizinho mais próximo (Nearest Neighbor)
- Florestas aleatórias (Random Forests)

Serão apresentados breves resumos de cada um.

### 3.1.2 Regressão Linear

Método ou análise permite prever o valor de uma variável em relação ao valor de outra, ou seja, a variável independente é a usada para prever o valor de outra e a que se deseja prever, é chamada de dependente, determinando, assim, os coeficientes de uma equação linear (IBM).

A regressão linear tenta minimizar as discrepâncias entre os valores de saída previstos, ajustando-se a uma linha reta. Esta análise é utilizada em diversas áreas como ciências biológicas, comportamentais, ambientais, sociais e em negócios.

Os modelos são simples, de fácil interpretação para gerar previsões fidedignas e podem ser treinados de forma rápida. ( $y = 1,5x + 2$        $y = ax + b$ )

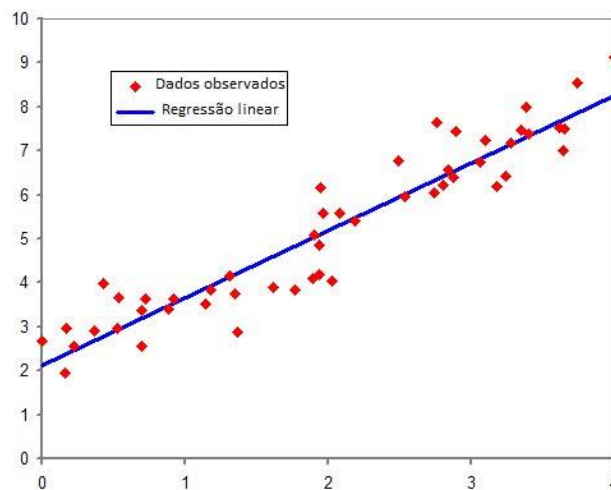


Figura 9 - REGRESSÃO LINEAR



### 3.1.3 Regressão Logística

A regressão logística, utilizada para análises preditivas, é uma ferramenta estatística. Muito utilizada no Marketing, Propaganda e Internet, para calcular riscos financeiros e seguros, na assistência médica, na detecção de fraudes etc., com o interesse de mensurar a probabilidade de um evento ocorrer.

Este modelo se aplica quando a variável dependente é binária (0 e 1) e utiliza a curva logística que representam a relação entre as variáveis.

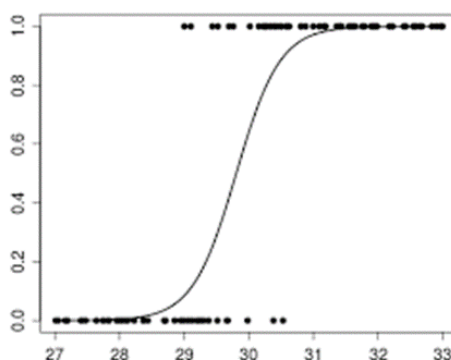


Figura 10 - REGRESSÃO LOGÍSTICA

### 3.1.4 Árvore de Decisão

A Árvore de Decisão é um algoritmo que divide o conjunto de dados de forma muito simples. Toma como entrada um vetor de valores e retorna uma decisão. Esta função divide um problema complexo em um problema mais simples, de forma recursiva. A árvore é composta por nós e folhas. O nó se divide em seus sucessores enquanto a folha é o último elemento. Os nós se dividem por teste condicional que representam decisões (BERTOZZO, 2019).

Este método é muito utilizado tanto em tarefas de classificação como de regressão. Quando as árvores são pouco profundas podem-se visualizar e interpretar facilmente. Tem como vantagem, requerer pouco esforço na preparação dos dados, pois geralmente não requerem normalização dos dados e conseguem lidar com valores faltantes, categóricos e numéricos.

### Árvore de Decisão para Jogar Tênis

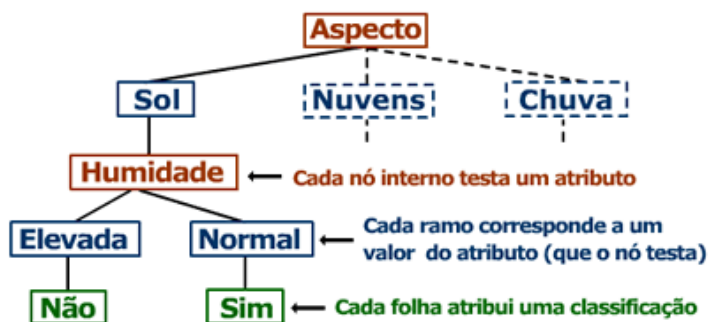


Figura 11 - ÁRVORE DE DECISÃO

#### 3.1.5 Máquinas de vetor de suporte

As Máquinas de vetor de suporte ou support vector machines SVMs conseguem resolver problemas de classificação e regressão. Durante a etapa de treinamento, adquirem a capacidade de generalização por meio do aprendizado. O objetivo é separar, por meio de uma função obtida pelos exemplos conhecidos na fase de treinamento, separar as instancias de duas classes num problema binário. Desta forma, será produzido um classificador que sirva para os dados não conhecidos, e assim prever saídas futuras. Este modelo gera um hiperplano de separação das duas classes, determinado pela distância das instâncias mais próximas (OLIVEIRA JUNIOR, 2010).

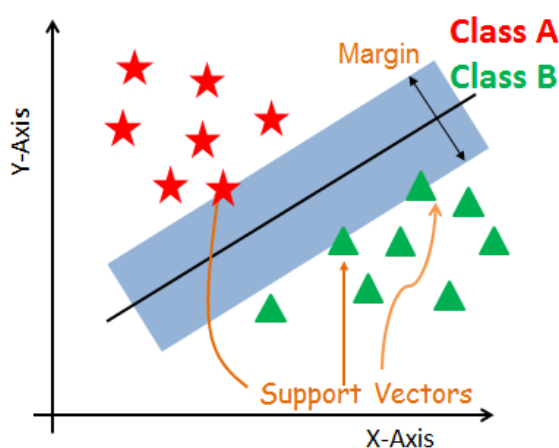


Figura 12 - MÁQUINAS DE VECTOR DE SUPORTE

### 3.1.6 Classificador Naïve Bayes

Este método é baseado no teorema de Bayes que descreve a probabilidade de um evento que se baseia num conhecimento anterior, relacionado ao evento e tem uma suposição “Ingênua” de independência condicional entre os pares de características.

O Teorema de Bayes tem esse nome pois, o pastor e matemático inglês Thomas Bayes foi o primeiro a fornecer uma equação que permitirá, que novas evidências modificassem a probabilidade de um evento a partir de um conhecimento anterior. A equação matemática do teorema é:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)} \quad P(B) \neq 0 \quad A \text{ e } B \text{ são eventos}$$

Este método, apesar de ser considerado de fazer suposições simplificadas e os classificadores serem ingênuos, funciona muito bem com classificação de documentos e filtragem de spam e pode ser muito mais rápido, comparado a métodos mais sofisticados.

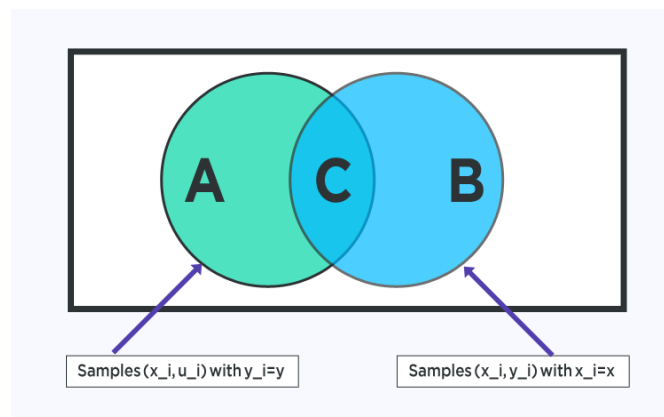


Figura 13 - CLASSIFICADOR NAYVE BAYES

### 3.1.7 Vizinho mais próximo

Este método não possui processamento na fase de treinamento. Ele serve para classificação e muito utilizado para reconhecimento de faces depois da extração das características. Utiliza-se a distância entre cada padrão de teste e todos os padrões de treinamento. Possibilita boas taxas de acerto principalmente se

o conjunto de treinamento é grande o inclui todas as variações possíveis dos dados. A principal desvantagem é a complexidade do teste.

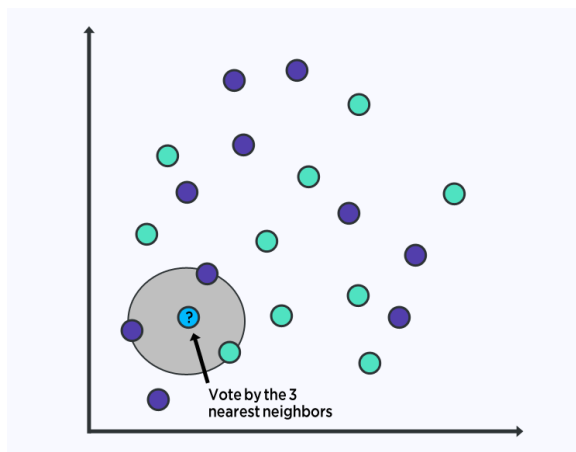


Figura 14 - VIZINHO MAIS PRÓXIMO

### 3.1.8 Florestas aleatórias

É um dos algoritmos mais utilizados. Ele é preciso, simples e flexível. Pode ser usado para tarefas de regressão ou classificação. Se adapta muito bem por não ser linear. Chama-se floresta porque está composta por uma floresta de árvores de decisão onde os dados são mesclados para ter previsões mais precisas e tem o benefício da aleatoriedade. Permite criar modelos com ampla diversidade e sem overfitting.

Pode ser usado para classificar um conjunto de dados relacionados a vinho, separando suas categorias, para identificar se um cliente vai deixar a empresa ou tentar prever quais clientes terão os maiores gastos em um ano.

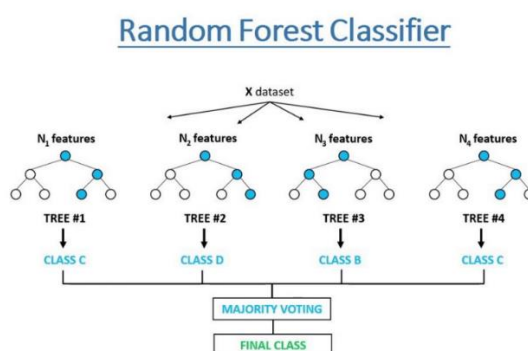


Figura 15 - FLORESTAS ALEATÓRIAS

## 3.2 Aprendizado Não Supervisionado

O objetivo principal do Aprendizado Não Supervisionado é encontrar padrões nos dados que não possuem rótulos, só se possui recursos. O modelo encontra seus próprios padrões. Geralmente é utilizado para detecção de anomalias e dividem os dados em grupos com base na similaridade. Pode-se ver um exemplo de uso em pacientes com doenças cardíacas e que precisam de diferentes tratamentos dependendo das características de cada um. Também é utilizado para observar áudios, imagens e vídeos, selecionar candidatos para uma vaga, separar turmas de alunos para um determinado trabalho etc.

Existem duas técnicas: de agrupamento ou clusterização e de associação. Estas técnicas podem ser aplicadas durante a mineração de dados, ou seja, na fase de pré-processamento. Ambas as técnicas serão abordadas resumidamente.

### 3.2.1 Técnica de Agrupamento ou Clusterização

Como dito anteriormente esta técnica não requer conhecimento prévio sobre as classes ou categorias dos dados. O conjunto de dados é segmentado em vários grupos ou clusters, de acordo com as semelhanças encontradas. Permite dividir um conjunto de dados em grupos referentes as medidas de distância ou de similaridade, aplicando-se fórmulas para se obter essas medidas.

Existem vários métodos que implementam esta técnica. Aqui só se fará uma menção deles. Os métodos são: Hierárquicos dividido em Agrupamento Aglomerativo e Agrupamento por Divisão, e os Não Hierárquicos dividido em K-MÉDIAS e K-MEDOIDES.

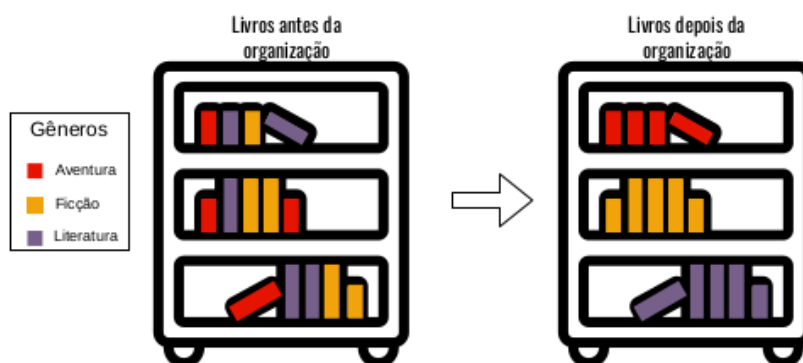


Figura 16 - TÉCNICA DE AGRUPAMENTO

### 3.2.2 Técnica de Associação

Esta técnica identifica conjuntos de itens que geralmente ocorrem juntos e assim consegue descobrir regras e correlações, ou seja, o objetivo da regra de associação é descobrir quais comportamentos acontecem com alta frequência no banco de dados e isto é chamado de Market Basket (IZBICKI e SANTOS, 2020). É muito utilizada para desenvolver estratégias eficazes de marketing e merchandising. Na figura a seguir, se evidencia a técnica de associação observando a relação que existe entre itens de uma cesta de mercado.

ID	Itens			
1	Pão	Leite		
2	Pão	Fralda	Cerveja	Ovos
3	Leite	Fralda	Cerveja	Refrigerante
4	Pão	Leite	Fralda	Cerveja
5	Pão	Leite	Fralda	Refrigerante

ID	Itens			
1	Pão	Leite		
2	Pão	Fralda	Cerveja	Ovos
3	Leite	Fralda	Cerveja	Refrigerante
4	Pão	Leite	Fralda	Cerveja
5	Pão	Leite	Fralda	Refrigerante

Figura 17 - TÉCNICA DE ASSOCIAÇÃO

Para montagem do modelo que permita a tomada de decisão, se utilizam três índices para realizar os cálculos. São estes:

- Suporte, que representa o percentual de vezes que um comportamento ou transação ocorre
- Confiança, que representa o percentual de confirmação de uma regra, sendo esta regra, toda combinação possível entre os comportamentos ou transações.
- Lift, que serve como validação da Confiança, apontando quantas vezes ocorre as combinações medidas no cálculo de Confiança

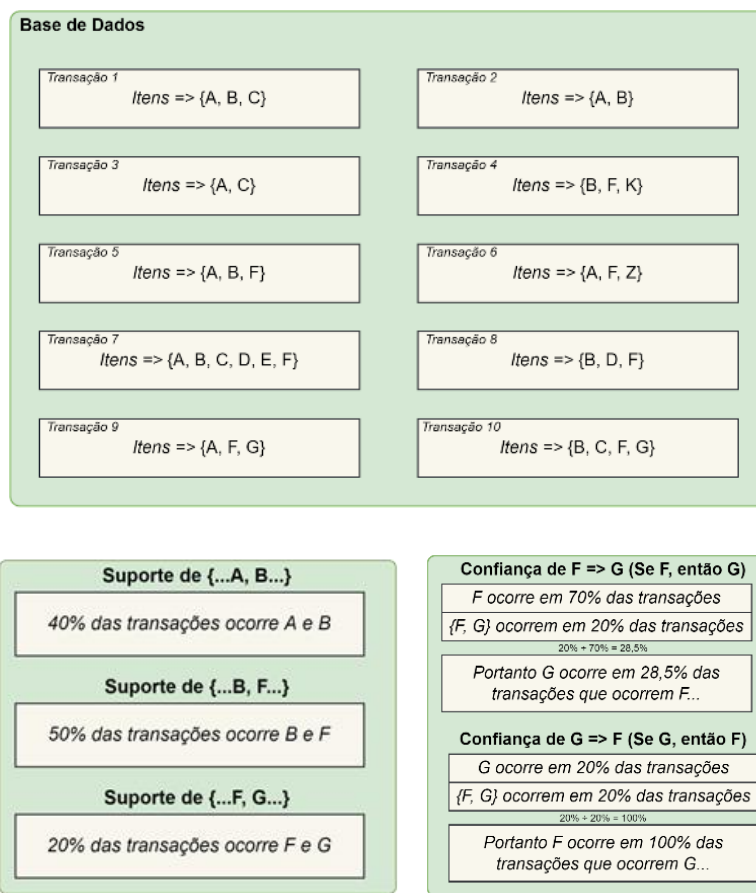


Figura 18 - SUPORT CONFIANÇA LIFT

Outra forma de cálculo é o Algoritmo APRIORI, que calcula os índices de todas as regras possíveis. Os elementos que não cumpram com o suporte mínimo, são eliminados e este comportamento proporciona um benefício, economizando processamento. O funcionamento ocorre da seguinte forma:

- Busca-se transações e seus itens.
- Checa-se a frequência de cada item.
- Caso atenda ao suporte mínimo, será adicionado à lista de “candidatos”.
- O processo se repete, checando-se a frequência dos “supersets” dos itens na lista de candidatos, gerando uma nova lista de candidatos até que a Lista (N) seja nula ou não tenha itens.

O gráfico a seguir ilustra um caso em que o item “a” não cumpre o suporte mínimo, portanto, é eliminado pelo algoritmo.

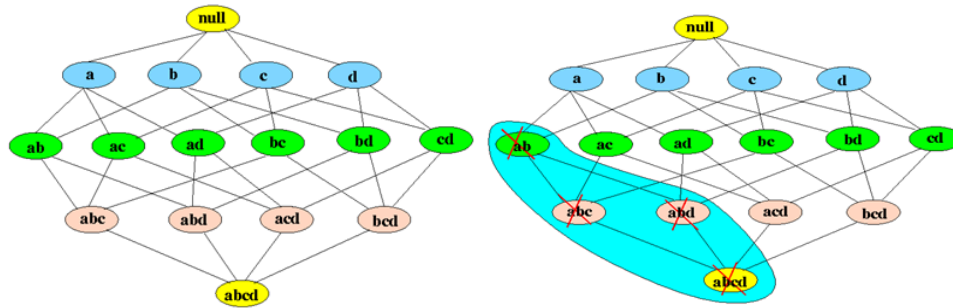


Figura 19 - ALGORITMO APRIORI

### 3.3 Aprendizado por reforço ou Reinforcement Learning

Esta categoria de aprendizado tem sua base em treinamento de modelos de Machine Learning para tomar uma sequência de decisões. É muito usada em Games, Robótica e é a principal técnica por trás do AlphaGo. (“O Que é Aprendizagem Por Reforço? - Deep Learning Book”)

O ambiente em que o agente precisa atingir uma meta, é incerto e complexo. Neste contexto, o sistema de inteligência artificial deverá enfrentar uma situação. Para encontrar uma solução a um problema, a máquina utiliza tentativa e erro a assim que ela consegue realizar o que o programador deseja, a inteligência artificial recebe recompensas, caso contrário, recebe penalidades. (“Uso da inteligência artificial para otimização do ... - PEBMED”) O aprendizado é iterativo e o objetivo da IA é maximizar a recompensa total.

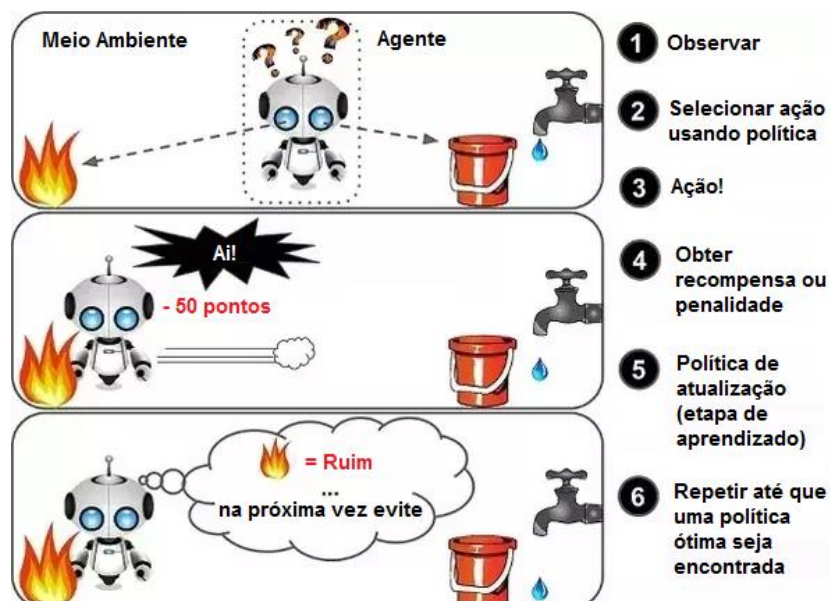


Figura 20 - APRENDIZADO POR REFORÇO



Apesar de ser o Cientista de Dados que determina as recompensas ou regras do jogo, em nenhum momento dará dicas ou sugestões ao modelo, para que este resolva o jogo. O processo se inicia com testes totalmente aleatórios e conclui com táticas sofisticadas.

Atualmente, este método é o mais eficaz de sugerir a criatividade da máquina, e é capaz de reunir experiência de milhares de jogos paralelos, se for executado em uma infraestrutura de computador poderosa.

Alguns algoritmos utilizados são:

- Q-Learning
- Aproximação por função com atualização de gradiente
- Multi-Armed Bandits
- Contextual Bandits e
- k-Armed Bandits

## 5. METODOLOGIA DA PESQUISA

O presente projeto se fundamentou em pesquisa bibliográfica e pesquisa por formulário digital. Diante da problemática deste trabalho, determinou-se que as técnicas de pesquisa a serem aplicadas sejam quantitativas e qualitativas.

Durante a pesquisa bibliográfica foram desenvolvidas diferentes etapas como a identificação e seleção do material bibliográfico pertinente. Posteriormente foi feita a leitura e o fichamento em formato digital e físico, do material selecionado, identificando obras, autores e suas ideias centrais, literatura nacional, internacional foi pesquisado no google acadêmico, no Scielo e algumas plataformas digitais. Por último foi feita uma análise do conteúdo do material levantado para poder ser utilizada na elaboração das conclusões da pesquisa (FERRAREZI JUNIOR, 2011).

Para desenvolver a pesquisa por formulário digital foram consideradas um conjunto de questões, sistematicamente articuladas, que se destinaram a levantar informações agrupadas, de forma digital, por parte das pessoas pesquisadas, de forma anônima, com o interesse de conhecer sobre o assunto em estudo. Foi elaborado, considerando-se os equipamentos, o tempo, as atividades, o ambiente e as condições dos indivíduos relativas à saúde. As questões realizaram-se em forma de fluxograma, tentando-se evitar erros. Todas elas, foram claramente formuladas para poder serem bem compreendidas pelos entrevistados, com questões objetivas, sucintas para evitar provocar dúvidas e ambiguidades (SEVERINO, 2010). Algumas, no formato fechado, por meio de múltipla escolha onde as respostas foram escolhidas dentre opções predefinidas pelo pesquisador, outras de forma aberta onde o entrevistado elaborou as respostas com suas próprias palavras, porém em textos curtos, com clareza da resposta solicitada.

O formulário digital foi aplicado a uma população específica: funcionários e estudantes de TI e que foi distribuído colaborativamente pelas redes sociais e pelos próprios entrevistados. O único pré-requisito foi pertencer ao segmento de tecnologia, definindo-se assim a amostragem da pesquisa.

Dada a natureza do tema estudado, gerou-se a necessidade de coletar e analisar dados já que o objetivo principal deste trabalho demanda evidências quantificáveis para seu melhor entendimento.

A traves de uma pesquisa quantitativa é possível mensural e quantificar as respostas dos entrevistados e assim obter dados que vão confirmar ou contestar a

hipótese inicial do trabalho, esperando-se que possibilite chegar à resolução do problema de pesquisa.

O procedimento de coleta de dados quantitativos foi aplicado com rigor para se obter a confiabilidade necessária para os resultados. Foram priorizados os resultados numéricos dos estudos propostos para avaliar as condições físicas e as condições de trabalho dos indivíduos.

O pesquisador tentou se limitar a descrição dos dados coletados aproveitando também os aspectos subjetivos que estão relacionados às respostas dos entrevistados uma vez que a interpretação e a análise dessas informações é tarefa deste trabalho.

A organização e apresentação dos dados levantados foram construídas em planilhas e demonstradas em forma de diferentes gráficos e tabelas, sendo esta, a última etapa do trabalho

## 6. RESULTADOS DA PESQUISA

Os dados obtidos com a pesquisa serão apresentados a seguir, conforme a análise realizada. Também foram incorporados elementos como tabelas e gráficos para que possam colaborar com a compreensão do leitor.

De acordo com Marconi e Lakatos (2003, p.231) a parte mais importante do trabalho é a interpretação dos resultados.

Foram coletados 384 formulários, respondidos em 20 países pertencentes ao continente americano e europeu.



Figura 21 - MAPA DE RESPOSTAS

Determinou-se no formulário, uma divisão em três categorias a seguir: Indivíduos que só trabalhavam, os que só estudavam e os que desenvolviam as duas atividades. A quantificação desses resultados, por país, se evidencia no gráfico a seguir.

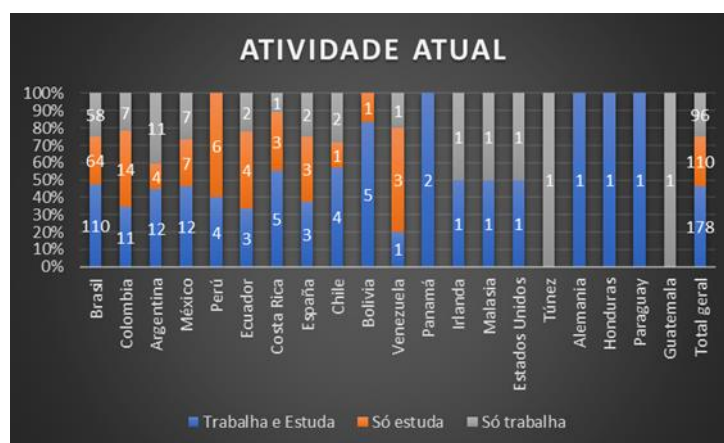


Figura 22 - ATIVIDADE POR PAÍS

Determinar a quantidade de respostas coletadas em função da atividade no total da amostra, é de suma importância para determinar a problemática.

Tabela 1 - ATIVIDADE ATUAL

	Atividade Atual
Trabalha e estuda	178
Só estuda	110
Só trabalha	96



Figura 23 - ATIVIDADE ATUAL

Os resultados de acordo com as respostas determinaram a seguinte classificação por sexo, sobre a população da amostra.

Tabela 2 - SEXO

Participantes	
Homem	267
Mulher	117
Total Geral	384

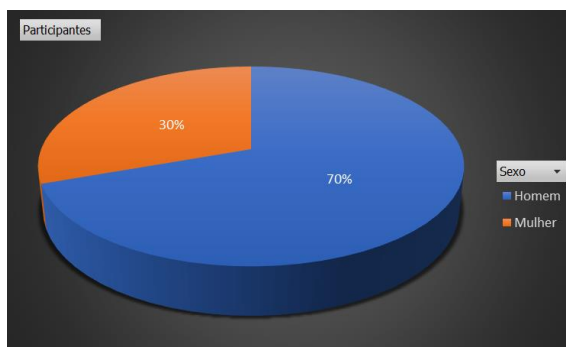


Figura 24 - SEXO

Devido à grande variedade das idades, foram divididas por frequência cada 10 anos.

Tabela 3 FREQUÊNCIA DE IDADE

Idade	Participantes
14 a 24	139
24 a 34	156
34 a 44	65
44 a 54	16
54 a 64	6
64 a 74	1
74 a 84	1

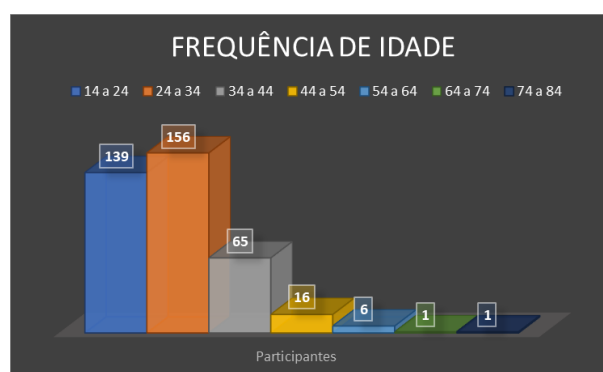


Figura 25 - FREQUÊNCIA DE IDADE

A seguir, é apresentada a relação dos indivíduos que descrevem dores nas costas, objeto do estudo, e as atividades que estes desenvolvem.

Tabela 4 - ATIVIDADE X DOR

	Dor nas costas		
	Sem resposta	Sim	Não
Trabalha e estuda	178		
Só estuda		72	38
Só trabalha	96		
Total general	274	72	38

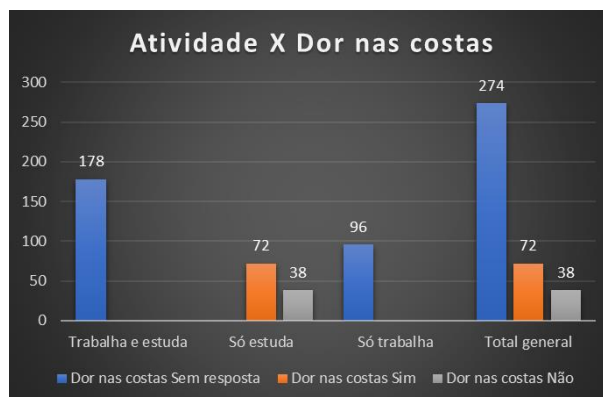


Figura 26 - ATIVIDADE X DOR

A prática de atividade física é um fator de suma importância para atrelar a falta de esta às dores nas costas, por esse motivo decidiu-se indagar ao respeito.

Tabela 5 - ATIVIDADE FÍSICA X ATIVIDADE ATUAL

Atividade	Pratica atividade física		
	Não	Sim	Total Geral
Só estuda	57	53	110
Só trabalha	42	54	96
Trabalha e estuda	89	89	178
Total Geral	188	196	384

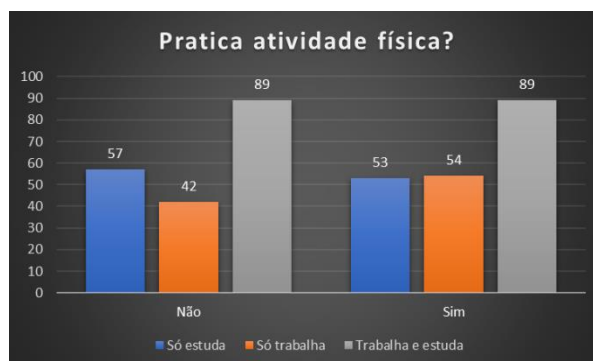


Figura 27 - ATIVIDADE FÍSICA X ATIVIDADE ATUAL

A permanência, por longos períodos, sentado frente ao computador é um fator para se ter em conta, devido aos prejuízos que isto ocasiona.

Tabela 6 - HORA/DIA/COMPUTADOR

Horas / Dia / Comp.	Participantes
De 0 a 4 horas	3
De 4 a 6 horas	32
De 6 a 8 horas	69
De 8 a 10 horas	156
Mas de 10 horas	124
Total Geral	384

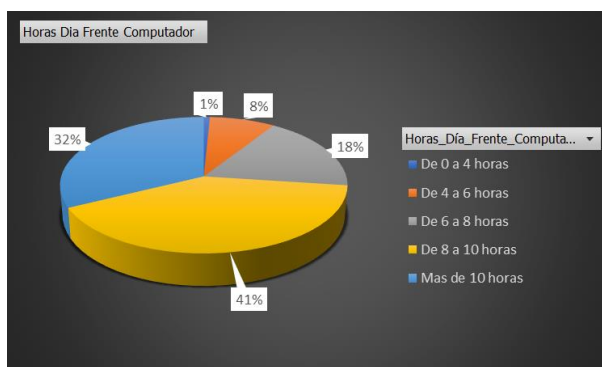


Figura 28 - HORA/DIA/COMPUTADOR

Saber quantas pessoas apresentam dores nas costas dentro da amostra coletada é o que a figura a seguir constata.

Tabela 7 - DOR NAS COSTAS

Dor nas costas	Participantes
Não	143
Sim	241
Total Geral	384

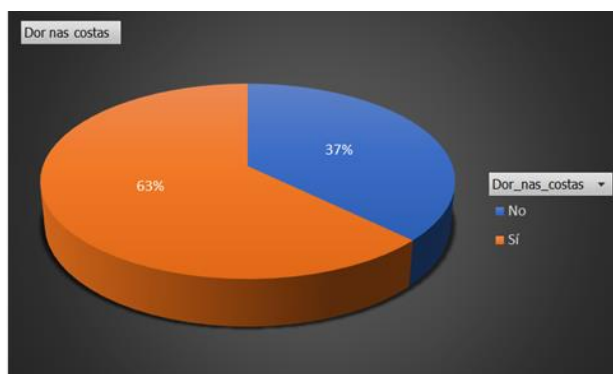


Figura 29 - DOR NAS COSTAS



As duas tabelas e seus respectivos gráficos a seguir, indagam a presença de dor nas costas antes e depois de fazer parte da categoria em que se baseia esta pesquisa.

Tabela 8 - DOR ANTES...

Dor antes...	Participantes
Sem resposta	131
Não	178
Sim	75
Total Geral	384



Figura 30 - DOR ANTES...

Tabela 9 - DOR DEPOIS...

Dor depois...	Participantes
Sem resposta	206
Não	84
Sim	94
Total Geral	384



Figura 31 - DOR DEPOIS...

Na próxima figura está representada a relação que existe entre as dores nas costas e a altura da tela do computador, na amostra coletada.

Tabela 10 - ALTURA TELA X DOR

Altura da tela	Dor nas costas		
	Não	Sim	Sem resposta
Mesma altura dos olhos	19	33	171
Abaixo da altura dos olhos	18	38	97
Por cima da altura dos olhos	1	1	6

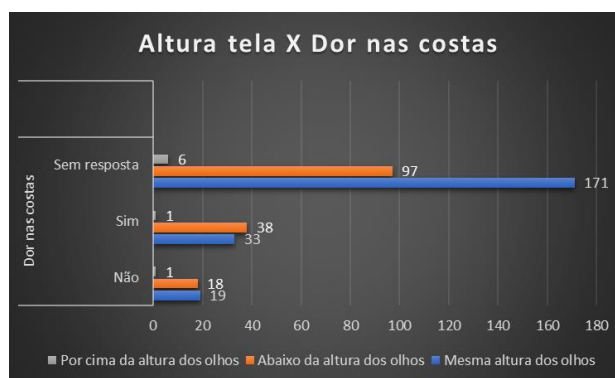


Figura 32 - ALTURA TELA X DOR

Relacionou-se também a incidência de dor com o tempo que o indivíduo passa sentado na frente do computador

Tabela 11 - HORAS/DIA X DOR

Horas / Dia / Computador	Dor nas costas		
	Sem resposta	Sim	Não
De 8 a 10 horas	123	22	11
Mas de 10 horas	95	19	10
De 6 a 8 horas	41	19	9
De 4 a 6 horas	12	12	8
De 0 a 4 horas	3		

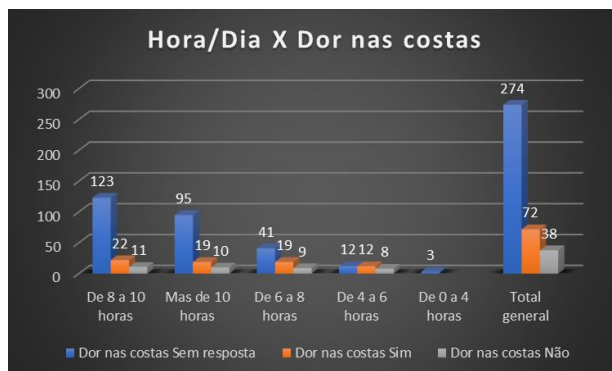


Figura 33 - HORAS/DIA X DOR

Foi segmentado exclusivamente, as pessoas que exercem a profissão.

Tabela 12 - HÁ QUANTOS ANOS TRABALHA?

Há quanto anos trabalha?	
De 1 a 3 anos	72
De 3 a 5 anos	23
Mas de 5 anos	88
Menos de 1 ano	91
Total Geral	274



Figura 34 - HÁ QUANTOS ANOS TRABALHA?

O gráfico a seguir tentou evidenciar a correlação de dor com o tempo que o indivíduo exerce a profissão.

Tabela 13 -TRABALHO X DOR

	Não	Sim	Sem dados	Total Geral
De 1 a 3 anos			72	72
De 3 a 5 anos			23	23
Mas de 5 anos			88	88
Menos de 1 ano			91	91
Não trabalham	38	72		110



Figura 35 - TRABALHO X DOR

É importante resaltar a preocupação dos empregadores com a condição física dos seus funcionários, antes das contratações.

Tabela 14 - SOLICITAR EXAMES

	A empresa solicitou exames?
Não	253
Sim	21
Sem resposta	110
Total Geral	384



Figura 36 - SOLICITAR EXAMES

Após o funcionário ser contratado, praticar ginástica laboral faz se necessário e as empresas deveriam proporcionar este benefício.

Tabela 15 - FAZ GINÁSTICA LABORAL

	Faz ginástica laboral?		
	Não	Sim	Total Geral
Só trabalha	87	9	96
Trabalha e estuda	164	14	178
Total Geral	251	23	274

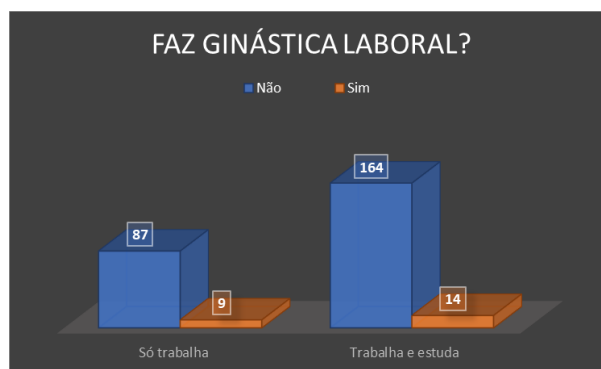


Figura 37 - FAZ GINÁSTICA LABORAL

Outras questões foram abordadas no formulário. São estas:

- Posição de uma segunda tela
- Distância do monitor
- Boa iluminação
- Intervalos de descanso
- Conhecimento sobre a ergonomia
- Uso de corretor de postura
- A dor nas costas limita o trabalho
- Notificar à empresa sobre as dores
- Atitude da empresa sobre a notificação
- Uso de acessórios para maior comodidade
- Tempo que sente dor nas costas
- Qual região das costas sente dor
- A dor causa estresse
- Visita ao médico para avaliação
- Exames após visita ao médico
- Diagnóstico após exames
- Necessidade de cirurgia
- Necessidade de fisioterapia
- Indicação de medicamento
- Tomar remédio por conta própria
- Modalidade de trabalho
- Equipamentos ergonômicos oferecidos pela empresa
- Medicamentos que usa

## 7. DISCUSSÃO SOBRE OS RESULTADOS DA PESQUISA

Um dos objetivos iniciais deste trabalho foi trazer a discussão sobre a possibilidade de utilização das técnicas de Data Mining através da coleta de dados por parte das empresas, para proporcionar uma melhor qualidade na saúde dos seus colaboradores, em relação aos distúrbios osteomusculares.

A motivação da pesquisadora se viu inspirada na própria percepção da dor, durante a pandemia, após passar períodos prolongados frente ao computador. Apesar desta, possuir conhecimentos sobre ergonomia e tentar adequar a mobília para evitar os transtornos e futuros comprometimentos da coluna, a dor foi inevitável. A partir desta situação, trouxe a inquietude de entender como os outros profissionais da área se encontrariam, principalmente após longos anos de exercício da função. Por este motivo, foi desenvolvido um formulário que conseguisse investigar de forma anónima, a maior quantidade de variáveis e condições possíveis para análise. Entendeu-se que o número de amostras era pequeno, contando com só 384 respostas em relação aos 30.000.000 (trinta milhões) de profissionais do setor no mundo.

No decorrer do trabalho, foram abordados os conceitos fundamentais de Data Mining e apresentada a pesquisa realizada nos profissionais da área de tecnologia, junto com seus resultados.

Com base nas informações levantadas e apresentadas na seção de resultados, verificou-se que houve divergências com os dados declarados pelos profissionais. Em algumas questões os indivíduos relataram não apresentarem dores relativas ao questionamento, porém em outros pontos da pesquisa, relatam que apresentam dores, o que gera uma informação pouco confiável para fazer uma análise.

Observa-se muitas abstenções no que se refere as dores após o início das atividades como profissionais, por outro lado, os trabalhos de pesquisa realizados por outros autores relacionados a este assunto, estão defasados e são muito escassos.

Apesar de ser um assunto muito relevante e que compromete a condição de saúde destes profissionais, pouco se fala do assunto.

Um fato interessante de ser abordado, é a falta de preocupação por parte das empresas, sobre as condições físicas previas ao ingresso do funcionário dado que uma atitude preventiva evitaria prejuízos para ambas das partes.

De acordo com a análise, a escassez dos dados não permitiu chegar a uma conclusão decisiva para comprovação da discussão presente. Acredita-se que a obtenção de um maior volume de dados possa contribuir para a alcançar resultados relevantes.



## 8. CONSIDERAÇÕES FINAIS

A presente pesquisa foi desenvolvida com o objetivo de investigar a existência de uma relação estreita entre o trabalho desenvolvido pelos profissionais de tecnologia e as dores nas costas. O resultado obtido não confirmou a hipótese.

Sabe-se que muitas das condições favoráveis para que o problema não ocorra, não são observadas pelas empresas. Pouco se estimula a prática de ginástica laboral e essa atividade seria de grande valia para a condição de saúde dos funcionários do setor. A ergonomia aplicada, os intervalos de descanso e outras medidas preventivas, preservariam a integridade deles.

Espera-se que este trabalho tenha contribuído para o desenvolvimento de um melhor entendimento da sociedade a respeito deste tema.

Devido ao crescimento do setor tecnológico e a consequente demanda de mão de obra qualificada, se faz necessária a busca da melhora da condição de saúde destes profissionais. Por este motivo, sugere-se então, novas e maiores pesquisas, indispensáveis para uma melhor compreensão dos fatos, já que contribui para a qualidade de vida das pessoas.

## REFERÊNCIAS

BERTOZZO, R. J. **Aplicação de Machine Learning em Dataset de consultas médicas do SUS**. 2019. 100f. Trabalho de Conclusão de Curso (Bacharel em Sistemas de Informação) – Universidade Federal de Santa Catarina, Santa Catarina, 2019.

FERRAREZI JUNIOR, C. **Guia do trabalho científico: do projeto à redação final: monografia, dissertação e tese**. São Paulo; Contexto, 2011.

IBM. IBM Cloud Education. Underfitting vs Overfitting. **IBM Cloud Learn Hub**, 23 mar. 2021. Disponível em: <https://www.ibm.com/cloud/learn/underfitting#toc-underfitti-FKaOZYJR> Acesso em: 24 jun. 2022, 22:57

IBM. IBM Cloud Education. Regressão linear. **IBM Cloud Learn Hub**, 23 mar. 2021. Disponível em: <https://www.ibm.com/br-pt/analytics/learn/linear-regression> Acesso em: 24 jun. 2022, 17:50

OLIVEIRA JUNIOR, G. M. **Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado**. 2010. 41f. Trabalho de Graduação (Bacharel em Ciência da Computação) – Universidade Federal de Pernambuco, Pernambuco, 2010.

Platzi: Cuso de Fundamentos Prácticos de Machine Learning - Educación. **Platzi**. 2022 180min. Disponível em: <https://platzi.com/cursos/fundamentos-ml/> Acesso em 5 maio 2022, 17:41

PROLABORE. Quais as profissões mais sujeitas a problemas na coluna? **Prolabore**, Rio Grande do Sul, 1 jun. 2022. Ergonomia. Disponível em: <https://pro-labore.com/quais-as-profissoes-mais-sujeitas-a-problemas-na-coluna/> Acesso em: 23 jun.2022, 10:35.

SEVERINO, A. J. **Metodologia do trabalho científico**. 23. ed. rev. e atualizada. São Paulo: Cortez, 2007.

SIDOW OSMAN, A. Data Mining Techniques: Review. **International Journal of Data Science Reserch**, Malaysia, 1 jun.2019. Disponível em: [https://www.academia.edu/77542877/Data\\_Mining\\_Techniques\\_Review](https://www.academia.edu/77542877/Data_Mining_Techniques_Review) Acesso em: 28 maio 2022, 18:56.