

Inteligência artificial para previsão de insuficiência cardíaca

Iuri Rodrigo Santos Matos – SENAI CIMATEC – iuri.matos@ba.estudante.senai.br

Rafael de França Moraes – SENAI CIMATEC – rafael.moraes@ba.estudante.senai.br

Vinícius José Pinto Dias – SENAI CIMATEC – vinicius.dias@ba.estudante.senai.br

Flavio Santos Conterato (orientador) – SENAI CIMATEC – flavio.conterato@ba.docente.senai.br

RESUMO

Inteligência artificial em conjunto com pesquisas médicas para detecção de ataques cardíacos. A insuficiência cardíaca é uma grande causa de mortes, sendo tratada como mal súbito, pois não há como o paciente saber quando irá sofrer ou se está sofrendo até que seja tarde demais. O objetivo desse artigo é ajudar as pessoas para que saibam se estão passando por esse problema, informando-as antecipadamente. Por meio de uma rede *multilayer perceptron* (MLP) utilizando sintomas e informações corpóreas de pacientes para treinar um modelo, o qual dará o devido suporte ao médico com base nas informações apresentadas no seu treino, se o hospitalizado analisado está sofrendo uma insuficiência cardíaca. O resultado do modelo tem 75% de acerto em detectar pacientes com a enfermidade. Assim mostrando ser capaz de na maioria dos casos onde paciente esteja sofrendo insuficiência seja detectado.

Palavras-chave: MLP. Matriz de Confusão. Insuficiência Cardíaca. Inteligência Artificial. Medicina.

1. INTRODUÇÃO

A insuficiência cardíaca ou infarto é uma das maiores causas de morte no Brasil e no mundo, onde segundo a OMS em território brasileiro são ceifadas 400 mil vidas por ano, já em escala mundial esse número se eleva para 17,5 milhões.

Como se é visto em produções cinematográficas ou em conhecimento popular, um dos sintomas do infarto é a dor forte no peito, porém há um problema nessa afirmação, a dor no peito é um dos sintomas mais comuns, porém ele não é o único e pode não acontecer em certos casos, isso eleva drasticamente a

mortalidade dessa condição onde não se sabe ao certo o que está acontecendo levando a uma busca tardia por um profissional da saúde.

A área da Inteligência Artificial está se expandindo para dar auxílio a outras diversas áreas do conhecimento, o que não seria diferente na medicina, como é dito por FERNANDO (2020), "Duas palavras prometem mudar a forma como você se relaciona com seu médico e ele se relaciona com você e os outros pacientes: inteligência artificial" e, por isso, esta pesquisa tem um intuito de dar um pequeno passo para o futuro próspero da fusão dessas áreas.

Tendo como resultado da pesquisa a predição de casos de insuficiência cardíaca por meio de um modelo *multilayer perceptron* (MLP), assim de modo a auxiliar para os próximos passos da Inteligência Artificial na área da medicina.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Insuficiência cardíaca (IC)

A insuficiência cardíaca é uma doença que assola pessoas de qualquer idade, segundo NOWELL (2020), "a insuficiência cardíaca é um distúrbio que atinge várias funcionalidades do coração, fazendo com que ele não consiga suprir as necessidades do nosso corpo", ela também pode variar em vários tipos como, por exemplo insuficiência cardíaca com fração de ejeção reduzida, insuficiência cardíaca com fração de ejeção preservada e entre outras variações.

Por ser uma doença que lenta e insidiosa seus sintomas não são evidentes desde o início e eles também podem surgir subitamente, e entre os sintomas temos a falta de ar, fadiga, edema nas pernas e incapacidade de realizar atividades de esforço. E em relação às suas causas NOWELL (2020) afirma que dentre as causas da insuficiência cardíaca estão os distúrbios que afetam diretamente o coração, outros que indiretamente o afetam, miocardite, toxinas, anemia, insuficiência renal e

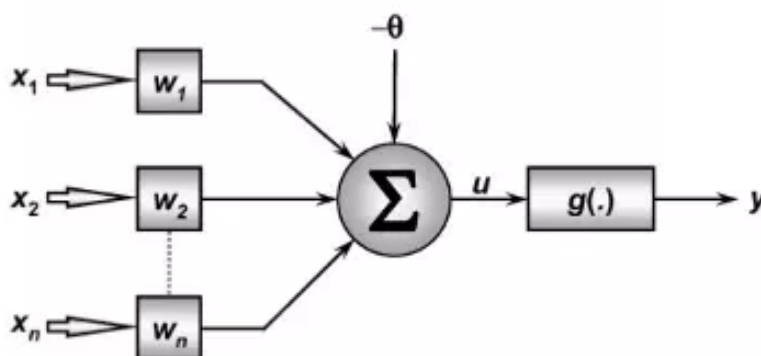
entre outros.

Mesmo sendo normalmente identificada pela aparição dos sintomas, alguns diagnósticos são feitos para ter a total certeza se o problema é realmente insuficiência cardíaca e pelo o que foi dito por NOWELL (2020) em sua matéria sobre a doença, entre os diagnósticos que confirmam a doença estão o radiografia do tórax, eletrocardiograma, ecocardiograma e exames de sangue. Para prevenir a IC basta tratar os distúrbios que o paciente sofre, como, por exemplo, obesidade, anemia, alcoolismo. E caso chegue na própria doença as maneiras de tratamento são dietas e mudanças no estilo de vida, tratamento da causa, medicamentos, etc.

2.2 Perceptron

O perceptron é uma rede neural com uma funcionalidade simples Segundo VINICIUS (2017), o mesmo é um classificador linear utilizado em treinamentos supervisionados, ele recebe uma entrada e a classifica em uma região específica ou não. Consistindo em quatro partes, primeiramente a rede recebe as entradas, que são multiplicadas por seus devidos pesos, logo após, os resultados são passados para a função agregadora que realiza as somas de todos os sinais, depois o resultado gerado por ela é somado ao Bias, que é considerado como mais um peso, e por fim é passado por uma função de ativação para gerar a saída certa.

Figura 1 - Estrutura de um perceptron



2.3 Multilayer Perceptron (MLP)

Uma MLP é um *perceptron* que trabalha com os conceitos de camadas ocultas e *backpropagation*. De acordo com LEONEL (2018) ele “pode ser pensado como uma rede neural artificial e profunda”.

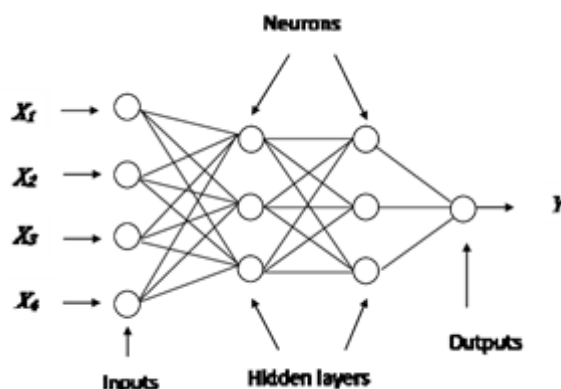
2.3.1 Backpropagation

Backpropagation é "um método popular para treinar redes neurais artificiais, especialmente redes neurais profundas". Onde, Segundo LEONEL (2018), ela é necessária para calcular o gradiente que é utilizado para a adaptação dos pesos das matrizes para assim minimizar a diferença do output real para o almejado.

2.3.2 Camadas ocultas

As camadas ocultas são nós ligados às entradas e saídas dos perceptrons, eles que permitem a possibilidade das redes neurais aprenderem coisas mais complicadas.

Figura 2 - Estrutura de um MLP



Sendo composto por mais de um perceptron, o MLP possui uma entrada, uma saída e as camadas ocultas entre elas, seu treinamento possui um objetivo bem simples, que é minimizar o erro o máximo possível, é aí onde entra a *backpropagation*.

Redes como a MLP atuam em constantes movimento de ida, se movendo da entrada para saída, e volta, onde a *backpropagation* e a regra da cadeia de cálculo são utilizadas, essa ação nos dá um gradiente onde os parâmetros podem ser ajustados até chegar ao mínimo de erro, objetivo que ao ser alcançado para essa ida e volta da rede.

2.4. Função de ativação

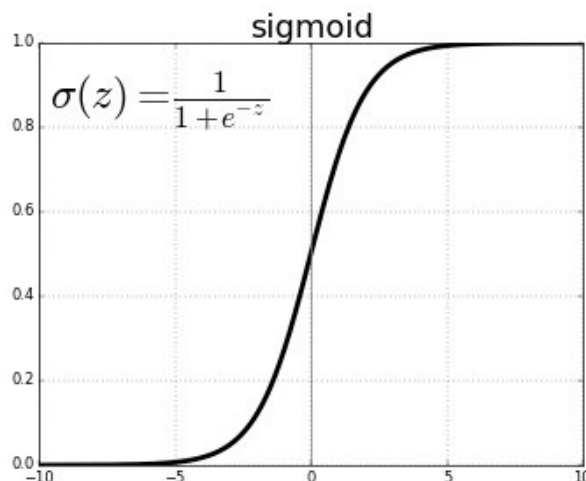
Por ser de um assunto que está em constante evolução, às funções de ativação são criadas para suprir as necessidades geradas, Segundo o *DEEP LEARNING BOOK* (2021), “Uma função de ativação é uma função matemática utilizada nas redes neurais, antes do output, para solucionar problemas mais inteligentes”. E por conta desse fator existem várias delas, dentre elas temos a sigmóide.

2.4.1. Sigmóide

Representada pela fórmula $f(x) = \frac{1}{(1 + e^{-x})}$, ela é uma função caracterizada por não gerar uma saída linear, mesmo que várias camadas utilizem a mesma como função de ativação, gerando um resultado que varia entre 0 e 1, essa função tende a colocar os dados de y nos extremos das saídas, gerando um característico formato de “S”.

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (2.4.1)$$

Figura 3 - Função de ativação sigmóide



Fonte: ichi.pro

2.5 Matriz de Confusão

Composta de 4 partes, uma matriz de confusão é uma matriz onde se é mostrado a frequência de classificação das classes do modelo, Segundo EMANUEL (2019) suas partes são o verdadeiro positivo, quando a classe buscada é prevista corretamente, o falso positivo, quando é previsto positivo incorretamente, o verdadeiro negativo, quando é previsto negativo corretamente, o falso negativo, quando o negativo é previsto incorretamente.

3. MATERIAIS E MÉTODOS

Esta pesquisa teve uma método de pesquisa com abordagem do tipo quantitativa, onde ocorre a tradução de dados obtidos em classificações e análises, natureza do tipo aplicada, pois, através dos dados estudados, geramos conhecimento para aplicações práticas para solucionar o problema que encontramos. Em relação ao objetivo, o tipo de pesquisa utilizada foi a descritiva onde caracterizamos o fenômeno da insuficiência cardíaca através da relação entre as variáveis dos dados presentes no *dataset* e quanto aos procedimentos o tipo de

pesquisa utilizada foi a experimental, onde selecionamos a variável que mais influencia no resultado desejado.

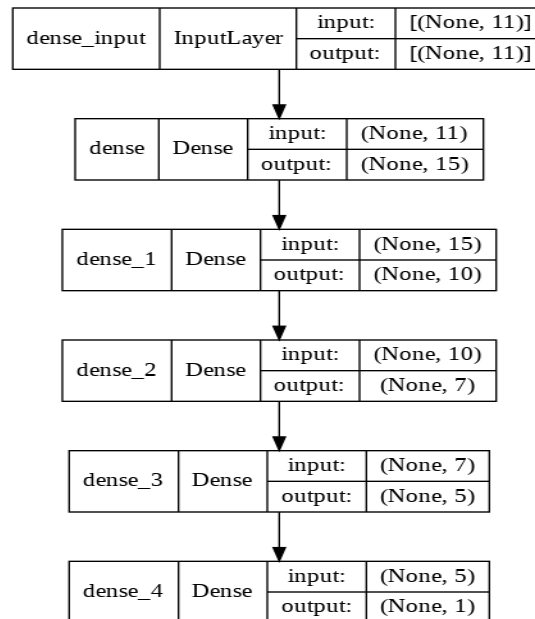
Todos os dados coletados para a pesquisa foram originados do *Heart Failure Prediction Dataset*, feito por FEDESORIANO (2021), focado para insuficiência cardíaca. O mesmo é uma combinação de 5 conjuntos de dados cardíacos diferentes, sendo eles Cleveland (303 observações), Húngaro (294 observações), Suíça (123 observações), Long Beach VA (200 observações) e Stalog (270 observações) Totalizando 1190 amostras, sendo 272 repetidas.

Todos esses conjuntos de dados são combinados em 12 características comuns, sendo elas idade, sexo, tipo de dor torácica, pressão arterial em repouso, colesterol, açúcar no sangue em jejum, resultados de eletrocardiograma em repouso, frequência cardíaca máxima alcançada, angina induzida por exercício, valor numérico medido na depressão (ST), inclinação do pico do segmento ST do exercício e doença cardíaca, sendo a última a que define se o paciente possui ou não a doença. Todas elas foram utilizadas para a geração do modelo.

Em relação aos parâmetros das camadas ocultas no modelo, ele contém um total de 463 parâmetros, divididos em 5 camadas densas contendo respectivamente 15, 10, 7, 5, 1 neurônios como também contém a função de ativação Sigmóide em todas as camadas.

Os parâmetros utilizados na compilação do modelo como perda, otimizador e métrica, são do próprio *Tensor Flow Keras*. Na função de perda foi utilizada a *BinaryCrossentropy* com argumentos padrões do *Keras*, utilizando a variável de perda como *binary cross-entropy*. O otimizador *Adamax* foi usado com os argumentos de ritmo de *learning rate* (aprendizado) em 0.004 e de resto por padrão. A métrica utilizada foi a *accuracy* (acurácia), sem nenhuma mudança nos argumentos padrões e utilizando o nome *accuracy* para a variável da métrica.

Figura 6 - Layout das camadas



Fonte: O Autor (2021)

4. RESULTADOS E DISCUSSÃO

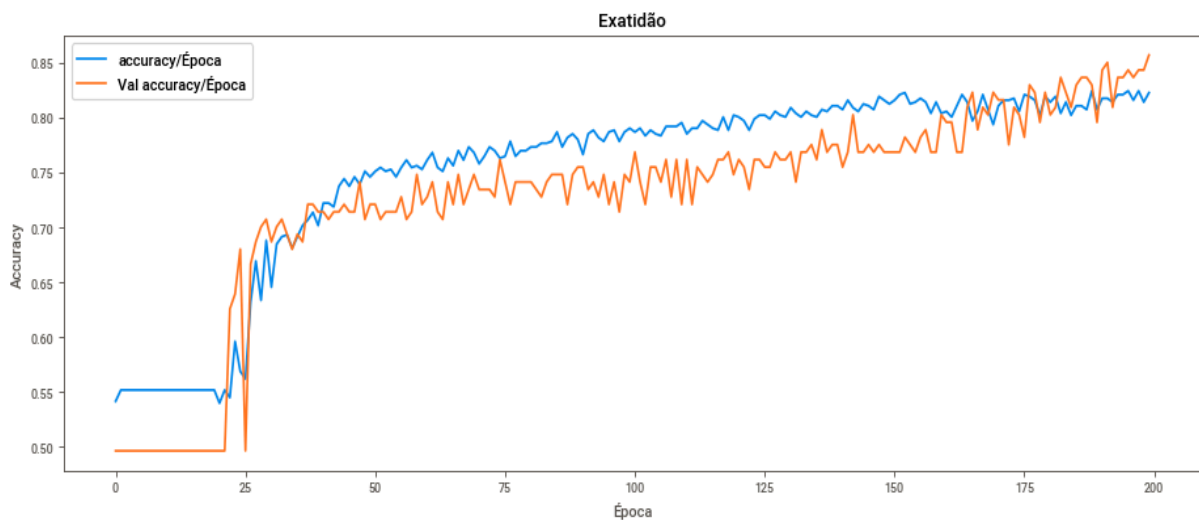
Os resultados do modelo mostrou que, através da utilização de dados de pacientes anteriores, onde foi utilizado o “*Heart Disease*” como dado de validação, é possível criar uma rede que com uma precisão de 75% que detecta e relata se um paciente está sofrendo uma insuficiência cardíaca, levando a um melhor cenário onde o paciente saberá com certeza o que está sofrendo, levando-o a um tratamento prematuro com melhores chances do quadro não se agravar.

Os resultados do modelo são medidos de diversas formas, essas formas nos ajudam a ver onde o modelo está pecando para que possamos resolver com melhor precisão o problema, dentre essas formas se encontram a acurácia, que nos dá a quantidade de acertos do modelo, o *loss* (perda) o qual traduzido do inglês significa perda, nos mostra exatamente o que o nome propõem, a quantidade de erro do nosso modelo, além de outras como a precisão onde é checado se em casos verdadeiros quantos realmente deram verdadeiro, levando para um cenário onde

falhas não são bem aceitas.

Todas as métricas de validação utilizadas se encontram nas figuras 7 a 9 , sendo eles de acurácia, perda e matriz de confusão.

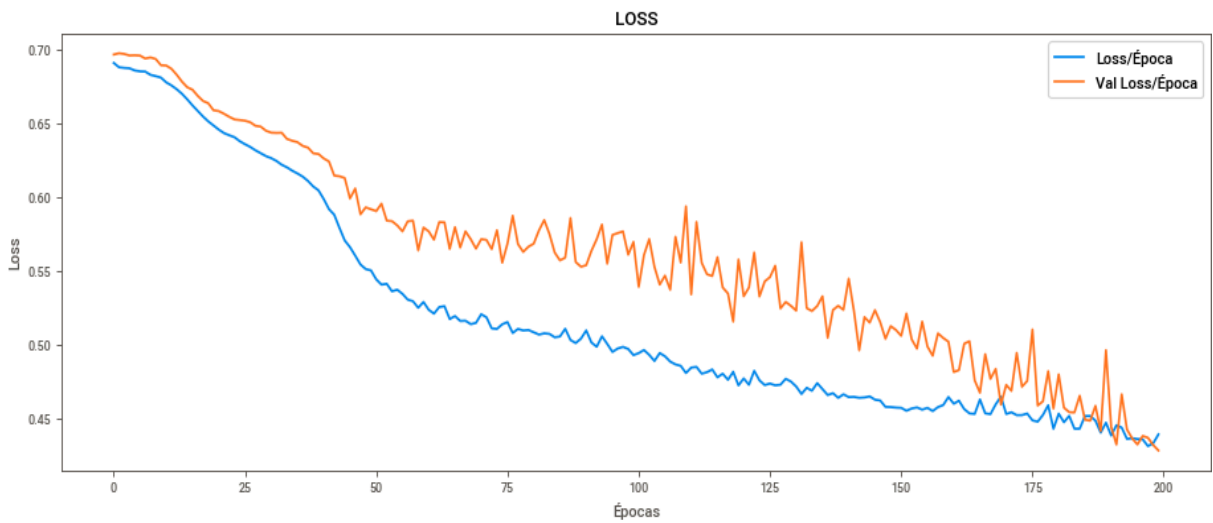
Figura 7 - Gráfico de Acurácia



Fonte: O Autor (2021)

A acurácia tende a manter um mesmo padrão nas épocas iniciais, tendo um aprendizado travado a um certo valor, o gráfico em questão é o resultado de 75% de acurácia.

Figura 8 - Gráfico de Perda

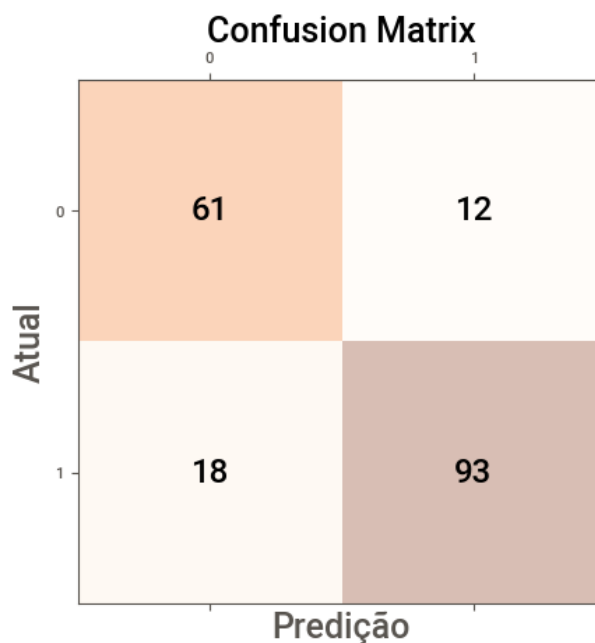


Fonte: O Autor (2021)

O gráfico de *loss* tende a ter variações assim como os valores de acurácia, variando de acordo com cada treino, nesse em questão, o valor de *loss* mostra um certo acompanhamento com o valor de *loss* por época.

Os dados utilizados na matriz são dados *Y (HeartDisease)*, contendo as instâncias com dados da insuficiência cardíaca, e dados preditos com o predict do modelo em dados normalizados do *StandardScaler*, contendo 20% total dos dados do dataset, sendo eles os 20% utilizados no teste e apenas os 80% pro treino.

Figura 9 - Matriz de Confusão



Fonte: O Autor (2021)

A matriz de confusão feita na figura 9 é a representação entre valores preditos e valores reais dos dados, onde 0 e 1 é onde não ocorre e ocorre insuficiência cardíaca, tendo a relação entre os dados preditos de casos com ocorrência de insuficiência cardíaca com os verdadeiros, onde tal matriz foi feita utilizando todos os dados de teste do modelo.

Com base nos gráficos e dados apresentados pode se observar que o modelo construído consegue dentro de uma boa margem de erro prever pacientes que estão sofrendo um caso de insuficiência cardíaca, concluindo o objetivo primário da pesquisa.

5. CONSIDERAÇÕES FINAIS

Por meio dessa pesquisa foi alcançado os objetivos estabelecidos no início da mesma, onde por meio de uma rede MLP (*multilayer perceptron*) foi realizado a predição em pacientes com insuficiência cardíaca.

Tendo resultados finais que poderiam chegar em um patamar melhor, esta pesquisa poderá servir de base para outras futuras que queiram melhorá-la ou utilizá-la de base para outros assuntos da área médica, área a qual ainda tem muitas pesquisas em conjunto com a área de Inteligência Artificial para serem realizadas.

Na qual esta pesquisa contribui diretamente no crescimento da área para outros âmbitos do conhecimento humano, servido como inspiração ou base para futuras descobertas e pesquisas, tendo seu pequeno porém muito importante papel na evolução da área em um todo.

6. REFERÊNCIAS

ANÁLISE DE REGRESSÃO LINEAR. **PRO Educacional**. Disponível em: <<https://proeducacional.com/ead/curso-cga-modulo-i/capitulos/capitulo-4/aulas/analise-de-regressao-linear/>>. Acesso em: 03 de Dezembro de 2021.

COELHO, Beatriz. Tipos de pesquisa: abordagem, natureza, objetivos e procedimentos. **Mettzer**, 2019. Disponível em: <<https://blog.mettzer.com/tipos-de-pesquisa/>>. Acesso em: 03 de Dezembro de 2021.

DEEP LEARNING BOOK. **Deep Learning Book**, 2021. Disponível em: <<https://www.deeplearningbook.com.br/funcao-de-ativacao/>>. Acesso em: 03 de Dezembro de 2021.

FEDESORIANO. Heart Failure Prediction Dataset. **Kaggle**, 2021. Disponível em: <<https://www.kaggle.com/fedesoriano/heart-failure-prediction>>. Acesso em: 03 de Dezembro de 2021.

FERNANDO, Hans. A inteligência artificial vai revolucionar a medicina que conhecemos. **Veja Saúde**, 2020. Disponível em: <<https://saude.abril.com.br/blog/com-a-palavra/a-inteligencia-artificial-vai-revolucionar-a-medicina-que-conhecemos/>>. Acesso em: 03 de Dezembro de 2021.

FINE, Nowell. Insuficiência cardíaca (IC). **Manual MSD versão saúde para a família**, 2020. Disponível em: <<https://www.msdmanuals.com/pt/casa/dist%C3%BArbios-do-cora%C3%A7%C3%A3o-e-dos-vasos-sangu%C3%ADneos/insufici%C3%A2ncia-card%C3%ADaca/insufici%C3%A2ncia-card%C3%ADaca-ic>>. Acesso em: 03 de Dezembro de 2021.

LEONEL, Jorge. Multilayer Perceptron. **Medium**, 2018. Disponível em: <<https://medium.com/@jorgesleonel/multilayer-perceptron-6c5db6a8dfa3>>. Acesso em: 03 de Dezembro de 2021.

LEONEL, Jorge. Backpropagation. **Medium**, 2018. Disponível em: <<https://medium.com/@jorgesleonel/backpropagation-cc81e9c772fd>>. Acesso em: 03 de Dezembro de 2021.

O QUE É REGRESSÃO LOGÍSTICA?. **Tibco**. Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-logistic-regression>>. Acesso em: 03 de Dezembro de 2021.

SANTOS, Virgilio Marques Dos. O que é Regressão Linear? Saiba como fazer a sua. **FM2S**, 2017. Disponível em:

<<https://www.fm2s.com.br/regressao-linear-economizar-milhoes/>>. Acesso em: 03 de Dezembro de 2021.

SOUZA, Emanuel. Entendendo o que é Matriz de Confusão com Python. **Medium**, 2019. Disponível em:

<<https://medium.com/data-hackers/entendendo-o-que-é-matriz-de-confusão-com-python-114e683ec509>>. Acesso em: 10 de Dezembro de 2021.

VINICIUS. Perceptron – redes neurais. **Monolito nimbus**, 2017. Disponível em: <<https://www.monolitonimbus.com.br/perceptron-redes-neurais/>>. Acesso em: 10 de Dezembro de 2021.