

A TECNOLOGIA E SAUDE UMA SIMBIOSE

Alex Sandro Zastrow Wickeinecki¹

Silvio Cesar Viegas²

RESUMO

Após uma pesquisa sobre o aumento dos custos na área da saúde que crescem em média 6% ao ano e a defasagem dos exames realizados no qual realiza-se métodos invasivos nos pacientes proponho no projeto o uso de uma inteligência artificial para auxiliar na detecção de tumores. No projeto foi usado como base de dados um arquivo csv, no qual constam atributos sobre os tumores e seu diagnóstico. Utilizando a ferramenta Jupyter Notebooks esses dados foram tratados e normalizados, após essa etapa utilizando algoritmo de inteligência artificial no qual foi escolhido o KNN, onde se classifica em grupos pela menor distância entre eles chegou-se a uma média de acerto de 75% no diagnóstico de tumores. A aplicação do projeto em um primeiro momento seria o funcionário coloca os dados dos tumores no software e a inteligência artificial passa a possibilidade de ser câncer maligno ou benigno, com essa resposta para auxiliar no diagnóstico o profissional da saúde decide sobre a realização de um exame mais invasivo. O baixo custo do projeto é muito atrativo pois qualquer computador pode rodar o software porque a parte de treinamento da inteligência artificial é feito em um computador mais robusto e após obter os pesos para a aplicação no algoritmo o sistema se torna muito versátil rodando praticamente em qualquer computador. Posteriormente pretende-se usar algoritmos mais robustos, como redes neurais, no qual pode se fazer o uso de imagens para diagnosticar cada vez mais precisamente e precocemente os tumores. Acredito que temos toda a tecnologia em nossas mãos para melhorar o mundo para todos, basta nos dedicarmos todos os dias em algo que amamos para realizar coisas incríveis.

Palavras-chave: Saúde, Simbiose, Algoritmos, Redes neurais, Inteligência artificial

¹ Acadêmico de Análise e Desenvolvimento de Sistemas da FAQI / FAQI / Gravataí / Rio Grande do Sul / Brasil / alex21387@hotmail.com

² Mestre / FAQI / Gravataí / Rio Grande do Sul / Brasil / silvio.viegas@qi.edu.br

1. INTRODUÇÃO

Os gastos com saúde estão crescendo rapidamente representando 10% do produto interno bruto (PIB) mundial. Um novo relatório da Organização Mundial da Saúde (OMS) revela uma rápida trajetória ascendente dos gastos com saúde em todo o mundo, o que é particularmente notável em países de baixa e média renda, onde eles crescem em média 6% ao ano. (Saúde, 2019)

Estes dados partem da ideologia incorreta de que se o paciente apresenta algum risco de doença o melhor será realizar uma série de procedimentos invasivos muitas vezes desnecessários, muitos dos exames que são realizados anualmente determinados pelos sistema de saúde vem sendo questionados, pois além de serem detectados com procedimentos mais simples ainda podem gerar outros problemas relacionados a sua execução.

Segundo a Agência Nacional de Saúde Suplementar (ANS) em determinados casos o uso de métodos invasivos para diagnostico em pacientes pode oferecer mais ricos do que benefícios.

Visando reduzir estes índices e os riscos relacionados a realização destes, propõem-se a difusão de uma inteligência artificial inovando os métodos existentes com o propósito de diagnósticos de baixo custo e risco.

1.1Tema

Os softwares que utilizam inteligência artificial utilizam uma grande quantidade de dados que através de métodos como o Machine Learning (aprendizado de máquina), usam os dados para compreender situações e ambientes onde o sistema aprende padrões por si próprio.

Partindo dessa ideologia utilizar Machine Learning para fazer classificação de tumores através de dados se torna promissor.

1.2 Delimitação do Tema

A primeira versão do software tem seu escopo em analisar um padrão que determine se o tumor de mama é maligno ou benigno através de dados de um arquivo CSV.

Logo o treinamento de Machine Learning vai contar apenas com esse arquivo para fazer a classificação.

1.3 Problema

É possível que um programa de inteligência artificial utilizando Machine Learning possa substituir os antigos métodos de diagnóstico de forma mais rápida e de menor custo?

1.4 Objetivo Geral

Desenvolver uma inteligência artificial para diagnóstico de câncer de mama de baixo custo.

1.5 Objetivos específicos

Analisar os critérios para o diagnóstico de tumores.

Averiguar os custos e a acessibilidade para o diagnóstico através de um programa inteligência artificial.

Averiguar a acuracidade do diagnóstico.

1.6 Justificativa

Nos últimos anos os gastos com saúde vêm aumentando tanto pelos seus métodos de diagnostico antigos como pela forma que eles são realizados, partindo deste ponto vemos que a saúde necessita de novos métodos para auxiliar neste processo.

O diagnóstico precoce do câncer de mama é a estratégia fundamental para o sucesso do seu prognóstico, no entanto de acordo com o Tribunal de Contas da União (TCU) o diagnóstico de confirmação do câncer vem sendo realizado com grau de estadiamento avançado, porque os métodos usados atualmente são invasivos, isto é, a biopsia acarretando local e preparo para a sua realização, com o uso de novas tecnologias esse processo seria minimizado (uniaio, 2019). Sendo assim a inteligência artificial usando de métodos de classificações de tumores fundamentadas em pesquisas realizadas de pacientes com a enfermidade auxiliará no diagnóstico.

2.FUNDAMENTAÇÃO TEÓRICA

Nessa seção será abordado os problemas das intervenções cirúrgicas e suas complicações o uso da tecnologia no diagnóstico e posteriormente os conceitos de inteligência artificial e da linguagem de programação utilizada.

2.1Intervenção cirúrgicas e suas complicações

A infecção hospitalar (IH) é o evento adverso que mais mata nos serviços de Saúde em todo o mundo, de acordo com a Organização Mundial da Saúde (OMS), sendo um risco enorme para a segurança do paciente. (MV Informática Nordeste Ltda, 2017). "O exame desnecessário não é inócuo, ele mata, porque leva a intervenções na maioria das vezes desnecessárias" (Ribeiro, 2011). Pois existem muitos fatores envolvidos como a suscetibilidade do paciente, como por exemplo, o sistema imunológico debilitando, se o médico precisar realizar um exame invasivo, quebrando a barreira natural de proteção do paciente, já seria impossível o índice zero de infecção". (MV Informática Nordeste Ltda, 2017).

Para fins de classificação epidemiológica, a infecção hospitalar é toda infecção adquirida durante a internação hospitalar (desde que não incubada previamente à internação) ou então relacionada a algum procedimento realizado no hospital (por exemplo, cirurgias), podendo manifestar-se inclusive após a alta.

A tecnologia entra como figura-chave na resposta ao problema, ao permitir um maior controle do ambiente hospitalar. "O uso de soluções de gestão sistematiza os processos, construindo indicadores e não deixando o hospital esquecer nenhum ponto estratégico no combate às IH, contribuindo para a segurança do paciente". (MV Informática Nordeste Ltda, 2017).

2.2 Tecnologia no diagnóstico

A aplicação da tecnologia à saúde está trazendo uma revolução nos processos de diagnósticos, com o uso dela os diagnósticos estão mais precisos e eficientes em relação aos métodos tradicionais.

As vantagens para essa revolução podem ser inúmeras, mas dentre elas uma chama a atenção: a diminuição do alto índice de diagnósticos errados por falha humana e consequências negativas por negligência com o exame clínico.

Uma pesquisa realizada pelo Journal of the American Medical Association em 2012, estima que entre 10% e 20% dos pacientes norte-americanos são vítimas de erros na identificação de doenças e 40 e 80 mil pessoas morrem pelos diagnósticos malfeitos. (Lucas, 2015).

Um fator importante a destacar é que com o uso da tecnologia, o diagnóstico de doenças está cada vez mais precoce trazendo uma redução nos custos do tratamento, diminuindo as chances de exames mais incisivos, e o aumento das chances de cura do tratamento.

2.3 Inteligência artificial

A definição de inteligência artificial está relacionada à capacidade das máquinas de pensarem como seres humanos - de terem o poder de aprender, raciocinar, perceber, deliberar e decidir de forma racional e inteligente. E não é um conceito novo. Em 1956, John McCarthy, um professor universitário, criou o termo para descrever um mundo em que as máquinas poderiam “resolver os tipos de problemas que hoje são reservados para humanos.” (Salesforce, 2019).

Basicamente inteligência artificial utiliza uma quantidade que pode ser pequena ou gigantesca de dados, onde esses dados podem ser imagens, medidas, valores ou uma informação e através da utilização de algoritmos se torna capaz de aprender, classificar, prever ou até mesmo descobrir algum padrão nesses dados ou informações.

A maior vantagem da Inteligência Artificial na medicina é, sem dúvidas, o auxílio na diagnose de patologias. Sabemos que nem todo caso é diagnosticado com facilidade, mas com essa tecnologia o processo se tornará consideravelmente mais tranquilo com análises muito mais seguras.

2.4 Linguagem de programação

Para o desenvolvimento de um software são utilizadas linguagens de programação que são a interação homem-máquina. Segundo Monteiro “Linguagem de Programação é uma linguagem escrita e formal que especifica um conjunto de instruções e regras usadas para gerar programas (software). Um software pode ser desenvolvido para rodar em um computador, dispositivo móvel ou em qualquer equipamento que permita sua execução. Existem várias linguagens e elas servem para muitos propósitos. Alguns óbvios, como criar um software, outros menos, como controlar um carro ou uma torradeira”.

2.4.1 Python

Python é uma linguagem de programação de alto nível, interpretada, de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991. Atualmente possui um modelo de desenvolvimento comunitário, aberto e gerenciado pela organização sem fins lucrativos Python Software Foundation. Apesar de várias partes da linguagem possuírem padrões e especificações formais, a linguagem como um todo não é formalmente especificada (Graciomar, 2016).

A linguagem de programação python foi escolhida por possuir uma série de bibliotecas de aprendizado computacional que vão auxiliar no desenvolvimento do projeto. A IDE que será utilizado para teste e a outra ferramenta será citado nas outras subseções.

2.4.2 Pycharm

O PyCharm é um IDE (ambiente de desenvolvimento integrado) utilizado para programação em Python, que possui diversos recursos extremamente úteis e que facilitam deveras as tarefas de desenvolvimento de software – principalmente quando

comparado com o IDE padrão do Python, o IDLE (Reis, 2016).

O PyCharm é desenvolvido pela empresa tcheca JetBrains, sendo escrito em Java e Python, e está disponível para vários sistemas operacionais, como Windows, Linux e OS X.

2.4.3 Jupyter Notebooks

Em 2014, Fernando Pérez, criou um subproduto do projeto IPython que chamou de projeto Jupyter. Jupyter Notebook é um ambiente computacional web, interativo para criação de documentos “Jupyter Notebooks”. O documento é um documento JSON com um esquema e contém uma lista ordenada de células que podem conter código, texto, fórmulas matemáticas, plotagens e imagens. A extensão dos notebooks é “ipynb”.

O Jupyter Notebook apresenta uma interface Web construída sobre algumas bibliotecas open-source, como o IPython, OMQ, Tornado, jQuery, Bootstrap e o MathJax. Ele pode conectar a núcleos de diferentes linguagens de programação. Além do Python, pode conectar-se a linguagens como o R, Julia, Ruby, Scala e Haskell. Atualmente, são suportadas mais de 40 linguagens de programação (Decisões, 2018).

3 Metodologia

Metodologia Científica é o estudo dos métodos que devem ser seguidos, em ordem cronológica, para que um pesquisador realize um trabalho acadêmico bem estruturado e de qualidade. (Dias, 2019)

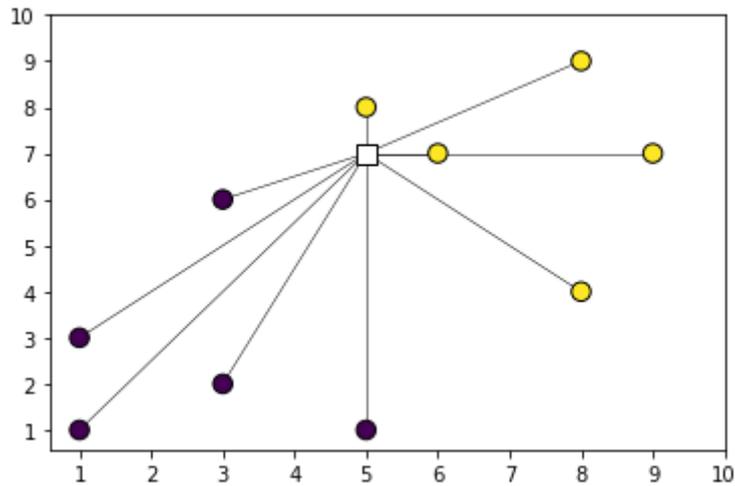
3.1 Pesquisa Aplicada

A pesquisa aplicada busca gerar conhecimento para a aplicação prática e dirigida a solução de problemas que contenham objetivos anteriormente definidos (Even3, 2018).

O projeto foi desenvolvido através da linguagem de programação python usando a IDE Pycharm e utilizando o ambiente computacional web, Jupyter Notebooks que auxilia na realização dos cálculos do algoritmo da IA.

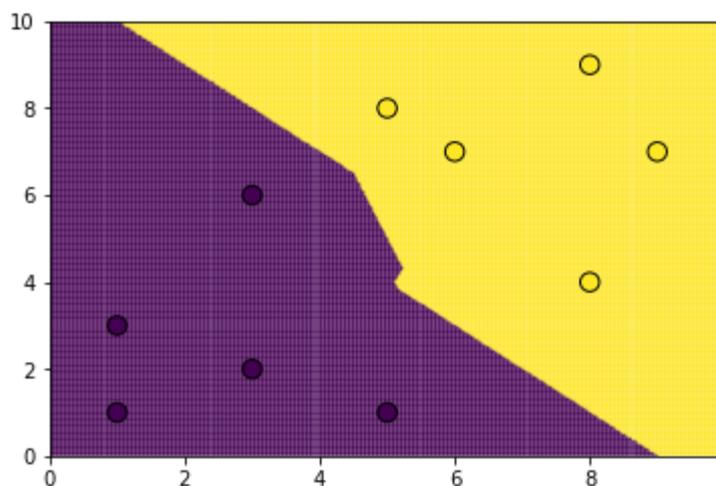
Os dados que foram utilizados no projeto são fornecidos pelo Instituto de Radiologia, Radiologia Ginecológica, Universidade Erlangen-Nuremberg, esses dados no formato de arquivo CSV contém 961 instancias de mamografias sendo que 516 delas foram consideradas benignas e 445 malignas. Os dados primeiramente foram tratados para evitar dados faltantes e valores fora da faixa, após esse procedimento foi realizada a normalização deles.

Para o desenvolvimento foi utilizado o paradigma supervisionado de classificação utilizando o algoritmo (KNN, do inglês k nearest neighbors) vizinhos mais similares, onde dado um conjunto de instâncias (i.e. treino), onde cada instância contém uma série de atributos e um deles é a classe, onde se pode fazer uma classificação através da distância euclidiana entre os pontos.



No gráfico acima se observa que o pequeno quadrado em branco está mais perto dos pontos amarelos, podendo ele ser classificado como amarelo por ter mais pontos amarelos próximos a ele.

No gráfico abaixo se visualiza a linha que divide os pontos amarelos dos roxos, qualquer ponto que for colocado em qualquer um dos dados vai ser classificado de acordo com sua posição ou seja. Se as coordenadas do ponto ficarem no lado amarelo esse ponto vai ser classificado como amarelo.



Após o tratamento e a normalização dos dados e da escolha do algoritmo de classificação, os dados foram divididos em partes sendo que um terço deles fica estabelecido como padrão para testes e os outros dois terços vão ser utilizados para o treino do algoritmo. O próximo procedimento é realizar o teste para ver se houve o

aprendizado utilizando os dados de testes.

Processado os testes se utiliza a matriz da confusão(A matriz de confusão é uma tabela que mostra as frequências classificação para cada classe do modelo), ou seja vai nos demonstrar quantas vezes o algoritmo acertou utilizando os dados de testes.

Após obter-se os melhores valores dos pesos do algoritmo(melhor acuracidade), eles são aplicados para a classificação no software.

A ideia inicial é ter uma interface com o operador no qual ele incorpora os dados da mamografia e a inteligência artificial identifica a probabilidade do tumor ser maligno ou benigno para que isso funcione corretamente o algoritmo deve ter a maior acuracidade possível

4.Desenvolvimento

Nessa primeira parte está sendo realizado a importação das bibliotecas e do arquivo CSV.

```
In [ ]: import pandas as pd
        from IPython.display import display
        mamografias = pd.read_csv('mamografias.csv')
```

Abaixo a descrição da base de dados, seus atributos e sua classe. Severity é a classe da base de dados ela informa o valor entre 0 e 1 para classificar os tumores sendo 1 para maligno e 0 para benigno.

Outra avaliação realizada nos dados é a diferença entre a média e a mediana onde se pode visualizar a presença de outliers e ruídos nos dados. Como os dois valores estão próximos é considerada a ausência de valores fora da faixa.

```
In [3]: mamografias.describe()
```

```
Out[3]:
```

	BI-RADS	Age	Shape	Margin	Density	Severity
count	830.000000	830.000000	830.000000	830.000000	830.000000	830.000000
mean	4.393976	55.781928	2.781928	2.813253	2.915663	0.485542
std	1.888371	14.671782	1.242361	1.567175	0.350936	0.500092
min	0.000000	18.000000	1.000000	1.000000	1.000000	0.000000
25%	4.000000	46.000000	2.000000	1.000000	3.000000	0.000000
50%	4.000000	57.000000	3.000000	3.000000	3.000000	0.000000
75%	5.000000	66.000000	4.000000	4.000000	3.000000	1.000000
max	55.000000	96.000000	4.000000	5.000000	4.000000	1.000000

```
In [6]: mamografias.median()
```

```
Out[6]: BI-RADS      4.0  
Age          57.0  
Shape        3.0  
Margin       3.0  
Density      3.0  
Severity     0.0  
dtype: float64
```

O próximo passo é descobrir o tipo de dado e a presença de algum valor nulo.

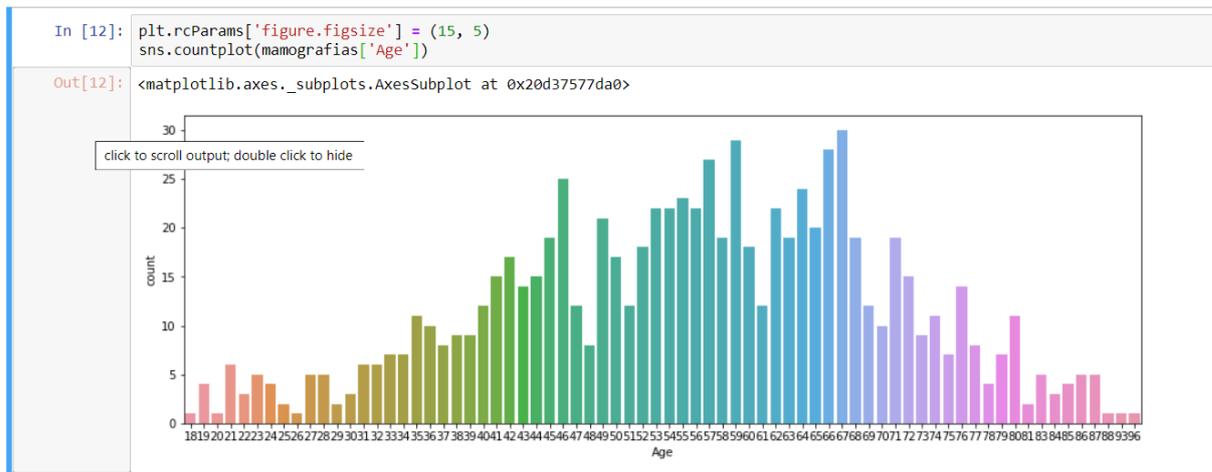
```
In [7]: mamografias.dtypes
```

```
Out[7]: BI-RADS      int64  
Age          int64  
Shape        int64  
Margin       int64  
Density      int64  
Severity     int64  
dtype: object
```

```
In [8]: mamografias.isnull().sum()
```

```
Out[8]: BI-RADS      0  
Age          0  
Shape        0  
Margin       0  
Density      0  
Severity     0  
dtype: int64
```

Uma das vantagens de utilizar a ferramenta Jupyter Notebooks é a facilidade de plotar gráficos que facilitam a visualização dos dados, o exemplo abaixo o número de amostras de mamografias de acordo com a idade.



A próxima etapa é a normalização dos dados ou seja deixar os valores dos atributos entre 0 e 1. Para fazer a normalização primeiro retiramos a coluna que contem a classe e colocamos a classe em outra variável.

```
In [11]: severidade=mamografias.pop("Severity")
```

```
In [12]: mamografias.describe()
```

```
Out[12]:
```

	BI-RADS	Age	Shape	Margin	Density
count	830.000000	830.000000	830.000000	830.000000	830.000000
mean	4.393976	55.781928	2.781928	2.813253	2.915663
std	1.888371	14.671782	1.242361	1.567175	0.350936
min	0.000000	18.000000	1.000000	1.000000	1.000000
25%	4.000000	46.000000	2.000000	1.000000	3.000000
50%	4.000000	57.000000	3.000000	3.000000	3.000000
75%	5.000000	66.000000	4.000000	4.000000	3.000000
max	55.000000	96.000000	4.000000	5.000000	4.000000

Após a exclusão realizamos a normalização.

```
In [18]: normalizada=(mamografias-mamografias.mean())/(mamografias.max()-mamografias.min())
normalizada.describe
```

```
Out[18]: <bound method NDFrame.describe of          BI-RADS      Age      Shape      Margin      Density
0      0.011019      0.143821      0.072691      0.546687      0.028112
1      0.011019      0.028437      0.406024      0.546687      0.028112
2     -0.007163     -0.356179     -0.593976     -0.453313      0.028112
3      0.011019      0.015616     -0.593976      0.546687      0.028112
4      0.011019      0.259206     -0.593976      0.296687      0.028112
5     -0.025345     -0.176691     -0.260643     -0.453313      0.028112
6     -0.007163     -0.253614      0.072691     -0.453313     -0.305221
7     -0.007163      0.054078     -0.260643     -0.453313     -0.305221
8     -0.007163     -0.022845     -0.593976     -0.453313      0.028112
9     -0.025345     -0.048486      0.072691      0.296687      0.028112
10    -0.007163      0.041257     -0.260643     -0.453313      0.028112
11    -0.007163     -0.022845     -0.593976     -0.453313      0.028112
12      0.011019      0.002796      0.406024      0.046687     -0.638554
13      0.011019     -0.176691      0.406024      0.296687      0.028112
14    -0.007163      0.041257     -0.260643      0.296687      0.028112
15      0.011019      0.246386      0.406024      0.546687      0.028112
16      0.011019     -0.138230      0.406024      0.546687      0.028112
17      0.011019     -0.010025      0.406024      0.296687      0.028112
18    -0.007163     -0.125409     -0.593976      0.546687     -0.305221
19      0.011019     -0.022845      0.406024      0.296687      0.028112
20      0.011019      0.015616      0.406024      0.296687      0.028112
21    -0.007163     -0.215153     -0.593976     -0.453313     -0.305221
22    -0.007163      0.323309     -0.593976     -0.453313      0.028112
23    -0.007163      0.054078     -0.260643     -0.453313      0.028112
24      0.011019      0.143821      0.072691      0.296687     -0.305221
25    -0.007163     -0.010025      0.072691      0.296687     -0.305221
26    -0.007163      0.284847     -0.593976     -0.453313     -0.638554
27    -0.007163     -0.074127     -0.593976     -0.453313      0.028112
28      0.011019      0.079719      0.072691      0.546687     -0.305221
29      0.011019      0.105360      0.406024      0.546687      0.028112
```

Após a normalização dos dados onde são reduzido aos limites [0,1] ou [-1,1], é inserida novamente a classe.

```
In [15]: normalizada.insert(loc=5, column='Severity',value=severidade)
```

```
In [16]: normalizada.describe
```

```
Out[16]: <bound method NDFrame.describe of          BI-RADS      Age      Shape      Margin      Density      Severity
0      0.011019      0.143821      0.072691      0.546687      0.028112          1
1      0.011019      0.028437      0.406024      0.546687      0.028112          1
2     -0.007163     -0.356179     -0.593976     -0.453313      0.028112          0
3      0.011019      0.015616     -0.593976      0.546687      0.028112          1
4      0.011019      0.259206     -0.593976      0.296687      0.028112          1
5     -0.025345     -0.176691     -0.260643     -0.453313      0.028112          1
6     -0.007163     -0.253614      0.072691     -0.453313     -0.305221          0
7     -0.007163      0.054078     -0.260643     -0.453313     -0.305221          0
8     -0.007163     -0.022845     -0.593976     -0.453313      0.028112          0
9     -0.025345     -0.048486      0.072691      0.296687      0.028112          0
10    -0.007163      0.041257     -0.260643     -0.453313      0.028112          1
11    -0.007163     -0.022845     -0.593976     -0.453313      0.028112          1
12      0.011019      0.002796      0.406024      0.046687     -0.638554          1
13      0.011019     -0.176691      0.406024      0.296687      0.028112          1
14    -0.007163      0.041257     -0.260643      0.296687      0.028112          1
15      0.011019      0.246386      0.406024      0.546687      0.028112          1
16      0.011019     -0.138230      0.406024      0.546687      0.028112          1
17      0.011019     -0.010025      0.406024      0.296687      0.028112          0
18    -0.007163     -0.125409     -0.593976      0.546687     -0.305221          0
19      0.011019     -0.022845      0.406024      0.296687      0.028112          1
20      0.011019      0.015616      0.406024      0.296687      0.028112          1
```

O comando a seguir demonstra quantas amostras da classe 0 e 1 temos no Data Frame.

```
In [17]: normalizada['Severity'].value_counts()
```

```
Out[17]: 0    427
         1    403
         Name: Severity, dtype: int64
```

Com a normalização dos dados e o tratamento dos mesmos é importado uma biblioteca chamada scikit-learn que inclui vários algoritmos de classificação, regressão e agrupamento. Através dessa biblioteca vamos aplicar o algoritmo KNN(K — Nearest Neighbors).

```
In [18]: from sklearn.model_selection import train_test_split
```

```
In [37]: X_train, X_test, y_train, y_test = train_test_split(normalizada.drop('Severity',axis=1), normalizada['Severity'], test_size=0.3)
```

```
In [38]: from sklearn.neighbors import KNeighborsClassifier
```

```
In [39]: knn = KNeighborsClassifier(n_neighbors=2)
```

```
In [40]: knn.fit(X_train, y_train)
```

```
Out[40]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                             metric_params=None, n_jobs=None, n_neighbors=2, p=2,
                             weights='uniform')
```

Na imagem acima foi importada a biblioteca, retirada novamente a classe 'Severity' do conjunto de teste, sendo ele correspondente a um terço do Data Frame e após isso foi importado o algoritmo KNN e aplicamos os dados de treino no algoritmo KNN.

Após a aplicação do algoritmo no conjunto de treino, é realizado no conjunto de teste.

```
In [41]: resultado = knn.predict(X_test)
         resultado
```

```
Out[41]: array([[1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1,
                1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0,
                0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0,
                0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0,
                0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1,
                0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1,
                0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0,
                0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
                0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0,
                0, 1, 0, 0, 0, 0, 1], dtype=int64)
```

Para prever os valores é utilizado o método predict() para os dados de teste. Para fazer a análise e a validação dos resultados foi utilizado a matriz de confusão, que é uma tabela que mostra as frequências de classificação para cada classe do modelo.

```
In [43]: print (pd.crosstab(y_test,resultado, rownames=['Valor Real'], colnames=[' 0 Predito'], margins=True))
```

Valor Real	0 Predito	1	All
0	115	14	129
1	48	72	120
All	163	86	249

A imagem acima é o resultados dos dados de teste passando pelo modelo. Dos 129 que pertencem a classe 0 o modelo acertou 115 e dos 120 que pertencem a classe 1 o modelo acertou 72 obtendo uma margem de acerto de 75,1%.

```
In [28]: a = float(input('Digite o valor do Bi-Rads :'))
b = float(input('Digite a idade :'))
c = float(input('Digite o valor do Shape :'))
d = float(input('Digite o valor da margin :'))
e = float(input('Digite o valor da densidade :'))

x_teste = np.array([[a,b,c,d,e]])
x_teste.shape
x_testenormalizada=(x_teste-x_teste.mean())/(x_teste.max()-x_teste.min())
diagnostico = knn.predict(x_testenormalizada)
if diagnostico == 1 :
    print('Severidade 1 : Maligno')
else :
    print('Severidade 0 : Benigno')
```

```
Digite o valor do Bi-Rads :5
Digite a idade :67
Digite o valor do Shape :3
Digite o valor da margin :5
Digite o valor da densidade :3
Severidade 1 : Maligno
```

Nessa última imagem foi colocado valores para os atributos e o algoritmo informou a resposta do possível diagnóstico.

5.Considerações finais

Os resultados obtidos com o projeto ao implementar algoritmos de machine learning em exames de mamografia foi satisfatório, pois a margem de acerto foi de 75,1%. Como o projeto está ligado a área da saúde uma margem de erro de 25% pode acarretar um grande número de diagnósticos falsos, pensando nisso em um primeiro momento a implementação do algoritmo ficaria sujeito a mudanças visando aumentar sua acuracidade e implementando algoritmos de deep learning e a utilização de

arquivos com imagens para fazer a classificação e até mesmo prever um possível diagnóstico.

REFERÊNCIAS

Bibliografia

uniao, t. d. (2019). Diagnóstico de câncer no Brasil é realizado de forma tardia. *Secom TCU*, <https://portal.tcu.gov.br/imprensa/noticias/diagnostico-de-cancer-no-brasil-e-realizado-de-forma-tardia.htm>.

Caldeira, H. (2017). A inteligência artificial aplicada na medicina. *cmtecnologia*, <https://cmtecnologia.com.br/blog/inteligencia-artificial/>.

Decisões, D. e. (2018). O JUPYTER NOTEBOOK: O QUE É? *Dados e Decisões*, <https://dadosedeciso.es.com.br/o-jupyter-notebook-o-que-e/>.

Dias, F. (2019). O que é Metodologia Científica e como utilizar no TCC. *Voitto*, <https://www.voitto.com.br/blog/artigo/o-que-e-metodologia-cientifica>.

Even3. (2018). Entenda a diferença entre a pesquisa aplicada e a pesquisa básica. *Even3*, <https://blog.even3.com.br/pesquisa-aplicada/>.

Graciomar. (2016). Como surgiu linguagem a Python. *Graciomar*, <http://www.graciomar.com.br/blog/10/como-surgiu-linguagem-a-python>.

Lucas. (2015). Tecnologia no diagnóstico – A revolução da saúde. *Cmtecnologia*, <https://cmtecnologia.com.br/blog/tecnologia-diagnostico-revolucioacao-saude/>.

Monteiro, L. P. (2015-2018). O que é linguagem de programação. *Universidade da tecnologia*, <https://universidadedatecnologia.com.br/o-que-e-linguagem-de-programacao/>.

MV Informática Nordeste Ltda. (2017). Infecção hospitalar: entenda o que é e suas principais causas no Brasil e no mundo. *MV Informática Nordeste Ltda*, <http://www.mv.com.br/pt/blog/infeccao-hospitalar--entenda-o-que-e-e-suas-principais-causas-no-brasil-e-no-mundo>.

Reis, F. d. (2016). Instalação do PyCharm, IDE para programação em Python. *Bóson treinamentos em Tecnologia*, <http://www.bosontreinamentos.com.br/programacao-em-python/instalacao-do-pycharm-ide-para-programacao-em-python/>.

Ribeiro, J. A. (2011). Cardiologista curitibano critica excesso de exames médicos e defende o retorno do médico generalista. *Revista Dimensão*, http://www.cruzvermelhapr.com.br/index.php?option=com_content&view=article&id=255%3Amenos-exames-e-mais-atencao-ao-paciente&catid=33%3Anoticias-gerais&lang=pt.

Salesforce. (2019). Entenda os Principais Conceitos e o que é Inteligência Artificial. *Salesforce Einstein*, <https://www.salesforce.com/br/products/einstein/ai-deep-dive/>.

Saúde, O. M. (2019). OMS: custos com saúde já representam 10% do PIB mundial. *onu news*, <https://news.un.org/pt/story/2019/02/1660781>.

Schulz-Wendtland, P. D. (2007). Mammographic Mass Data. *The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process.*, 4164-4172.

uniao, t. d. (2019). Diagnóstico de câncer no Brasil é realizado de forma tardia. *Secom TCU*, <https://portal.tcu.gov.br/imprensa/noticias/diagnostico-de-cancer-no-brasil-e-realizado-de-forma-tardia.htm>.