

Impacto de Fatores de Risco em Casos Confirmados de Covid-19 no Estado de São Paulo

Um Estudo com Regras de Associação e
Algoritmos para Classificação de Óbitos

Thalita Sylvia Marques

CONTEXTO

- ❖ Pandemia Covid-19 com mais de 350 milhões de casos confirmados ao redor do mundo
- ❖ Fenômeno *Infodemic*
- ❖ Necessidade em gerar conhecimento científico ou confiável sobre a pandemia



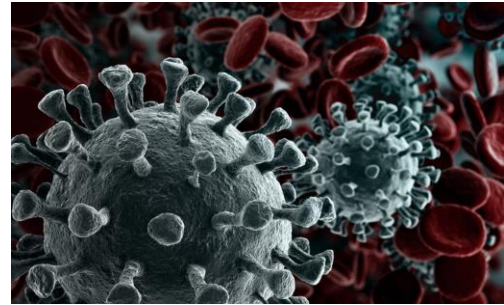
PROBLEMA

- ❖ Análise e interpretação dados
- ❖ Tomada de decisão
- ❖ Combate às *fake news*
- ❖ Estudar os impactos dos fatores de risco nos casos confirmados de Covid-19



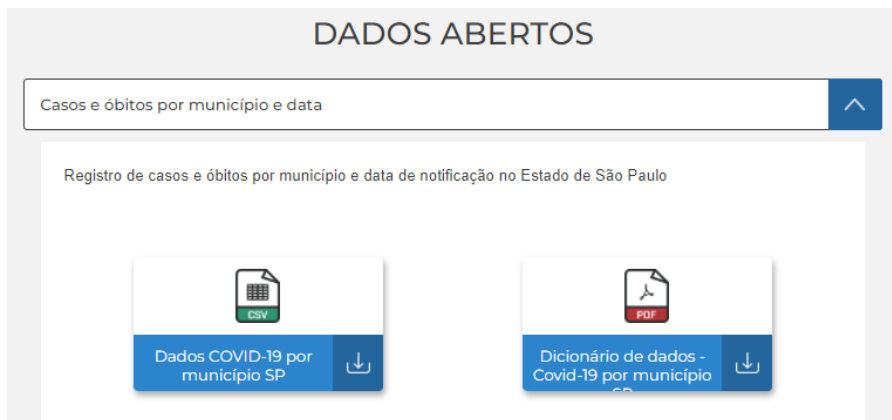
SOLUÇÃO

- ❖ Análise exploratória como Processo de Descoberta de Conhecimento
- ❖ Geração de modelos de Machine Learning com algoritmos para classificação
- ❖ Fatores de risco como preditores de óbito



OBTENÇÃO DOS DADOS

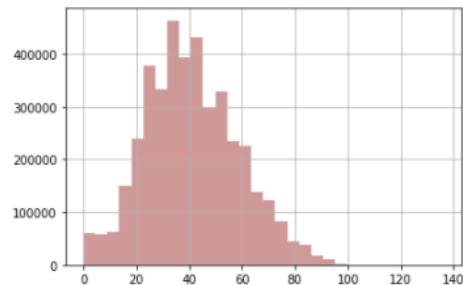
- ❖ Portal de dados abertos do Estado de São Paulo



ANÁLISE EXPLORATÓRIA

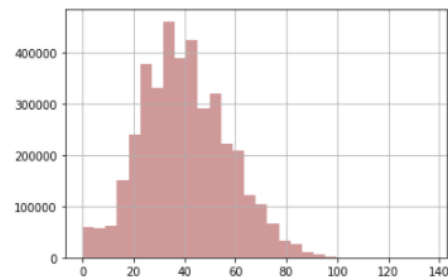
```
#Distribuição geral  
df['Idade'].hist(bins=30, color='darkred', alpha=0.4)
```

<AxesSubplot:>



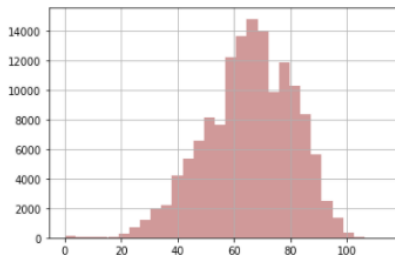
```
#Distribuição de idade entre os casos recuperados  
df[df['Obito']==0]['Idade'].hist(bins=30, color='darkred', alpha=0.4)
```

<AxesSubplot:>

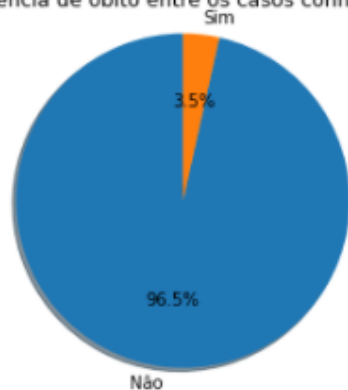


```
#Distribuição de idade entre os casos de óbito  
df[df['Obito']==1]['Idade'].hist(bins=30, color='darkred', alpha=0.4)
```

<AxesSubplot:>



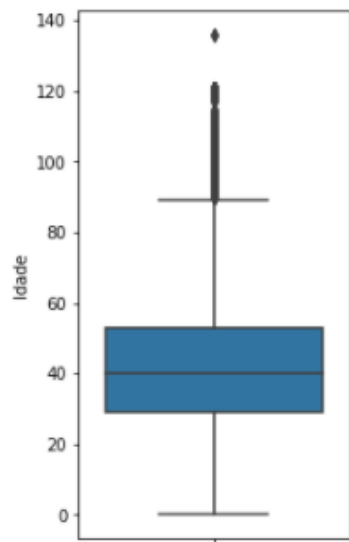
Ocorrência de óbito entre os casos confirmados



ANÁLISE EXPLORATÓRIA

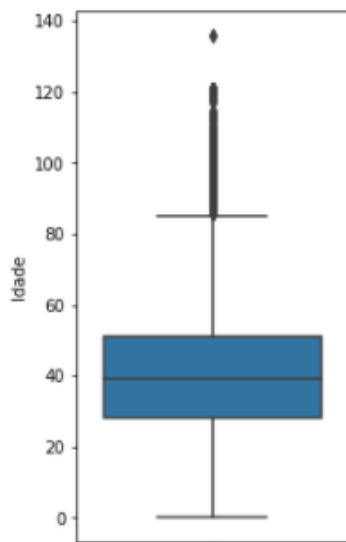
```
#Boxplot das idades (geral)
plt.figure(figsize=(3, 6))
sns.boxplot(y='Idade', data=df )
```

<AxesSubplot:ylabel='Idade'>



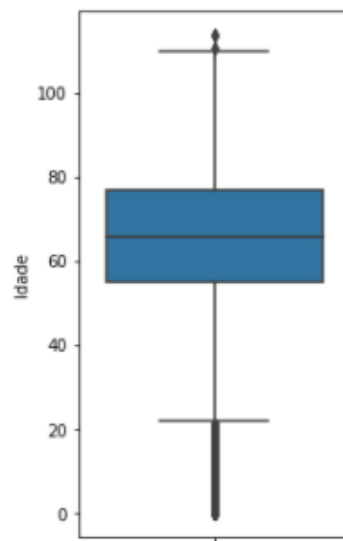
```
#Boxplot das idades (casos recuperados)
plt.figure(figsize=(3, 6))
sns.boxplot(y='Idade', data=df[df['Obito']=='=0'] )
```

<AxesSubplot:ylabel='Idade'>

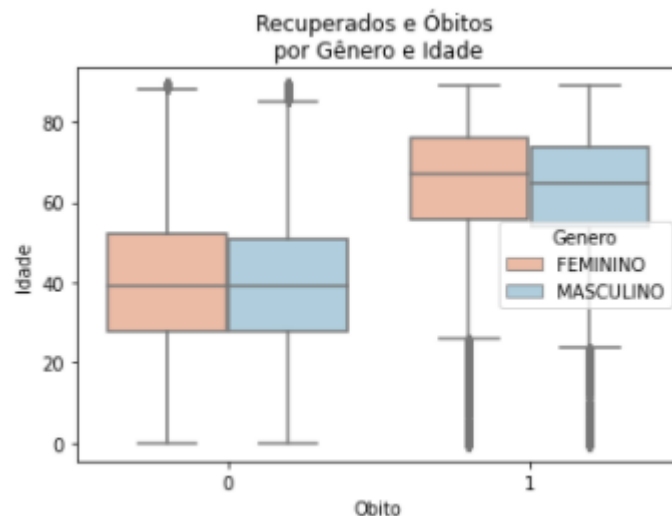
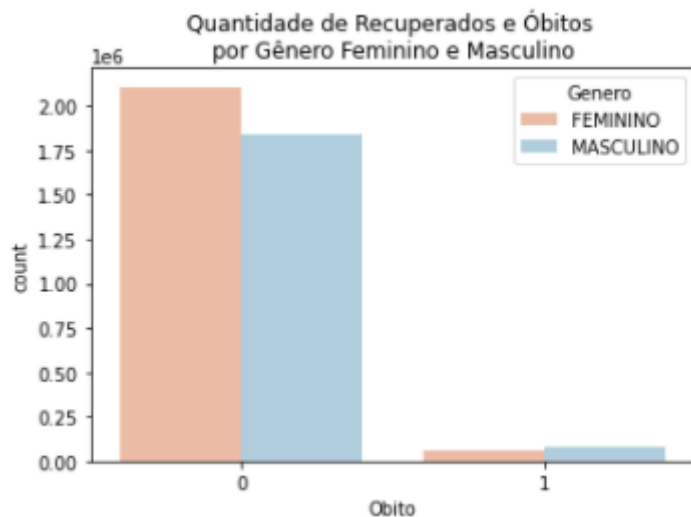


```
#Boxplot das idades (casos de óbitos)
plt.figure(figsize=(3, 6))
sns.boxplot(y='Idade', data=df[df['Obito']=='=1'] )
```

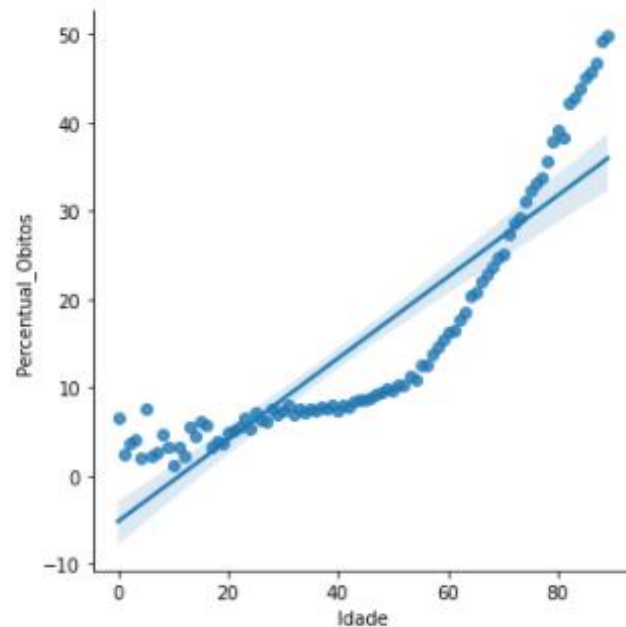
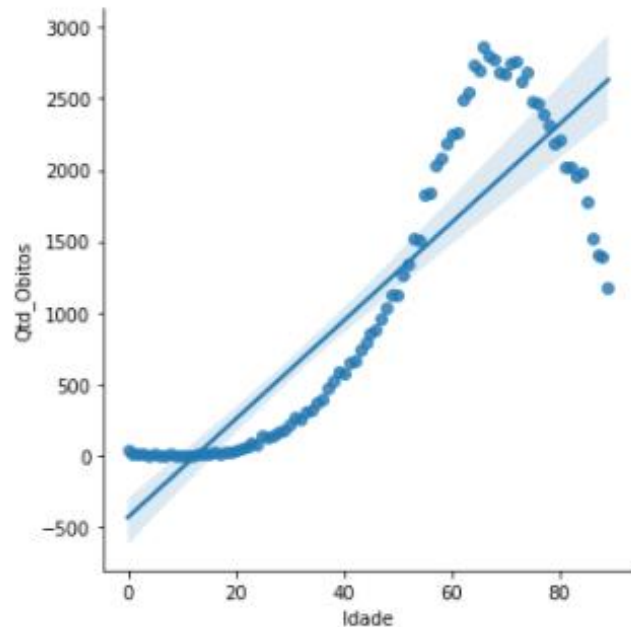
<AxesSubplot:ylabel='Idade'>



ANÁLISE EXPLORATÓRIA

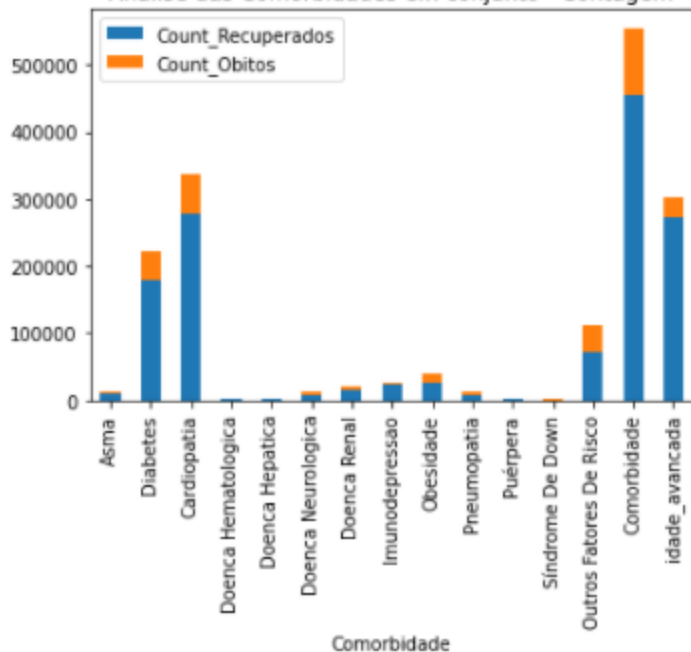


ANÁLISE EXPLORATÓRIA

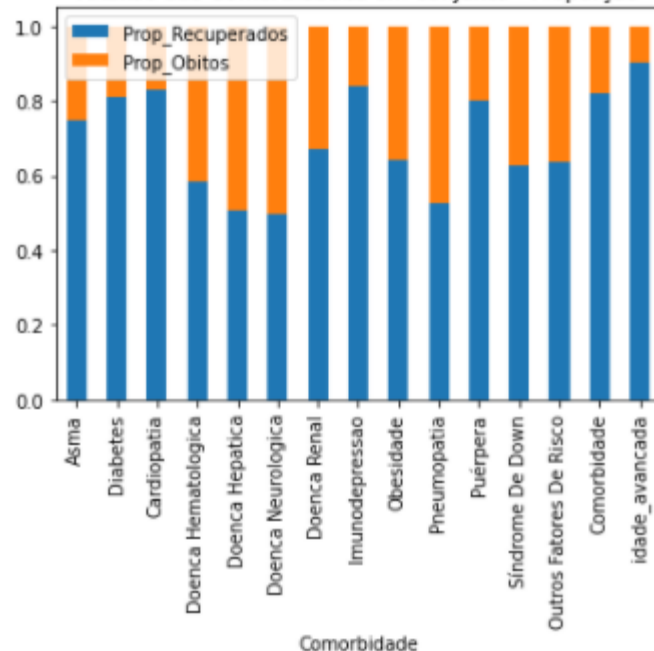


ANÁLISE EXPLORATÓRIA

Análise das Comorbidades em conjunto - Contagem



Análise das Comorbidades em conjunto - Proporção



REGRAS DE ASSOCIAÇÃO

	antecedents	consequents	support	confidence
0	(Doença Neurológica, idade_avancada)	(Obito)	0.010532	0.561906
1	(Outros Fatores De Risco, Genero_Masculino, idade_avancada, Cardiopatia)	(Obito)	0.012642	0.515982
2	(Outros Fatores De Risco, Diabetes, Genero_Masculino, idade_avancada)	(Obito)	0.010743	0.506666
3	(Doença Neurológica)	(Obito)	0.012518	0.503439
4	(Outros Fatores De Risco, Diabetes, idade_avancada, Cardiopatia)	(Obito)	0.010835	0.497602
5	(Obesidade, idade_avancada)	(Obito)	0.011267	0.493923
6	(Outros Fatores De Risco, Genero_Masculino, idade_avancada)	(Obito)	0.028214	0.487950
7	(Outros Fatores De Risco, idade_avancada, Cardiopatia)	(Obito)	0.023232	0.481457
8	(Pneumopatia)	(Obito)	0.011193	0.473965
9	(Outros Fatores De Risco, Diabetes, idade_avancada)	(Obito)	0.020270	0.471858
10	(Outros Fatores De Risco, Diabetes, Cardiopatia)	(Obito)	0.013497	0.461439
11	(Outros Fatores De Risco, Genero_Masculino, Cardiopatia)	(Obito)	0.016191	0.452934
12	(Outros Fatores De Risco, idade_avancada)	(Obito)	0.051494	0.451905
13	(Outros Fatores De Risco, Genero_Masculino, Diabetes)	(Obito)	0.014059	0.448278
14	(Diabetes, Obesidade)	(Obito)	0.010046	0.438714
15	(Outros Fatores De Risco, Cardiopatia)	(Obito)	0.029515	0.426361
16	(Outros Fatores De Risco, Diabetes)	(Obito)	0.026379	0.421434
17	(Obesidade, Cardiopatia)	(Obito)	0.013476	0.414796
18	(Outros Fatores De Risco, Genero_Masculino)	(Obito)	0.039734	0.385152

MACHINE LEARNING

- ❖ Foram desenvolvidos modelos de classificação com os seguintes algoritmos:
 - ❖ 1. Regressão Logística
 - ❖ 2. Árvores de Decisão
 - ❖ 3. Florestas Aleatórias
 - ❖ 4. Classificação Bayesiana

MACHINE LEARNING

- ❖ As tentativas de criar os modelos de classificação foram sendo registradas da seguinte forma no relatório:
 - ❖ 1. Tratamentos iniciais
 - ❖ 2. Faixas etárias
 - ❖ 3. Afunilamento do período
 - ❖ 4. Quantidade de comorbidades
 - ❖ 5. Undersampling
 - ❖ 6. Oversampling
 - ❖ 7. Hiperparâmetros (para alguns algoritmos)

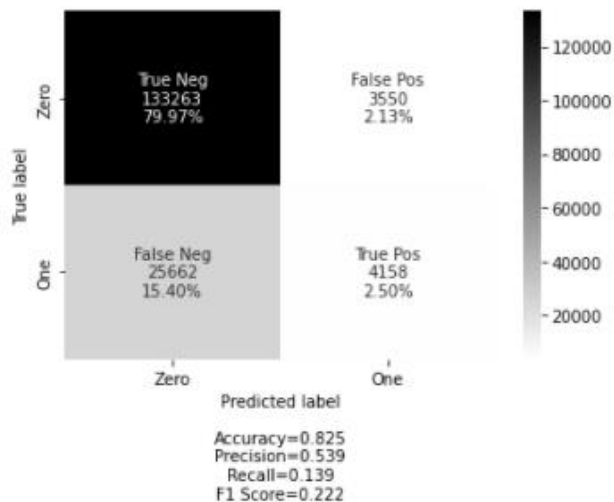
MACHINE LEARNING

- ❖ Os resultados dos modelos criados não se distanciaram muito
- ❖ Será destacada então a interpretação do primeiro modelo:
- ❖ **Regressão Logística** com os tratamentos iniciais

MACHINE LEARNING

Relatório de Classificação:

	precision	recall	f1-score	support
0	0.84	0.97	0.90	136813
1	0.54	0.14	0.22	29820
accuracy			0.82	166633
macro avg	0.69	0.56	0.56	166633
weighted avg	0.79	0.82	0.78	166633



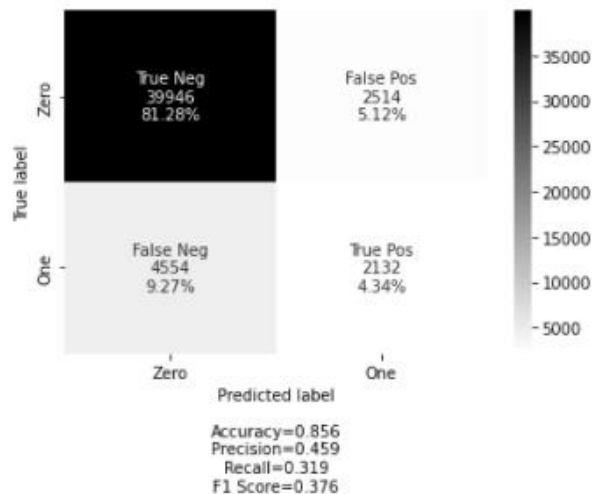
- ❖ Acurácia alta, porém próxima da precisão da classe zero
- ❖ A diferença entre o recall, ou revocação, para as classes 0 e 1, bem como os valores de precisão mostra que o algoritmo tem mais capacidade de acertar a classe 0
- ❖ A respeito do F1 Score, o baixo valor apresentado leva a interpretar que a acurácia provavelmente não é relevante e que os valores VP, VN, FP, FN podem apresentar distorções

RESULTADOS

- ❖ Modelo de maior valor de **precisão** para a classe de **recuperados**:
 - ❖ Obtido do algoritmo de Classificação Bayesiana, utilizando a tentativa número 4
 - ❖ Versão do tratamento dos dados com faixas etárias, período específico e coluna de quantidade de comorbidades

Relatório de Classificação:

	precision	recall	f1-score	support
0	0.90	0.94	0.92	42460
1	0.46	0.32	0.38	6686
accuracy			0.86	49146
macro avg	0.68	0.63	0.65	49146
weighted avg	0.84	0.86	0.84	49146

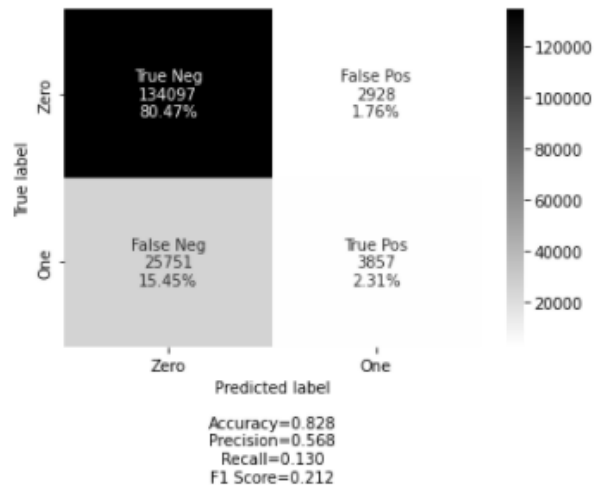


RESULTADOS

- ❖ Modelo de maior valor de **precisão** para a classe de **óbitos**:
 - ❖ Obtido com Árvores de Decisão e com Florestas Aleatórias com a primeira versão de tratamentos e resultados iguais

Relatório de Classificação:

	precision	recall	f1-score	support
0	0.84	0.98	0.90	137025
1	0.57	0.13	0.21	29608
accuracy			0.83	166633
macro avg	0.70	0.55	0.56	166633
weighted avg	0.79	0.83	0.78	166633



CONCLUSÃO

Este trabalho teve o objetivo de mostrar as comorbidades no desfecho de óbito dos indivíduos com Covid-19, com a intenção de gerar mais informação sobre o risco de óbito no caso de pessoas expostas a fatores de risco. Acredita-se que o objetivo foi concluído, mesmo que mais pela análise exploratória do que pelos modelos de Machine Learning.

OBRIGADA!