

Covid-19 - Prevendo o uso de leitos de UTI com Machine Learning

Elton Gil Sena





Introdução

- Dezembro de 2019 - primeiro caso oficial de Covid-19 é relatado;
- Fevereiro de 2020 - primeiro caso oficial de Covid-19 é registrado no Brasil;
- O aumento do número de casos faz a demanda por leitos hospitalares também aumentar;
- Medidas são tomadas para aumentar o número de leitos UTI;



Introdução

- Hospital Sírio Libanês libera uma base de dados no Kaggle com informações sobre pacientes com Covid-19;
- Inspirar a comunidade de Dados mundial para criar modelos de Machine Learning que possam prever a necessidade de um leito UTI por um paciente com Covid-19;



Objetivo

Desenvolver modelos de Machine Learning que possam realizar as tarefas:

1. Prever a necessidade de leitos UTI para pacientes com Covid-19;
2. Prever a NÃO necessidade de leitos UTI para pacientes com Covid-19;



Estrutura do Projeto

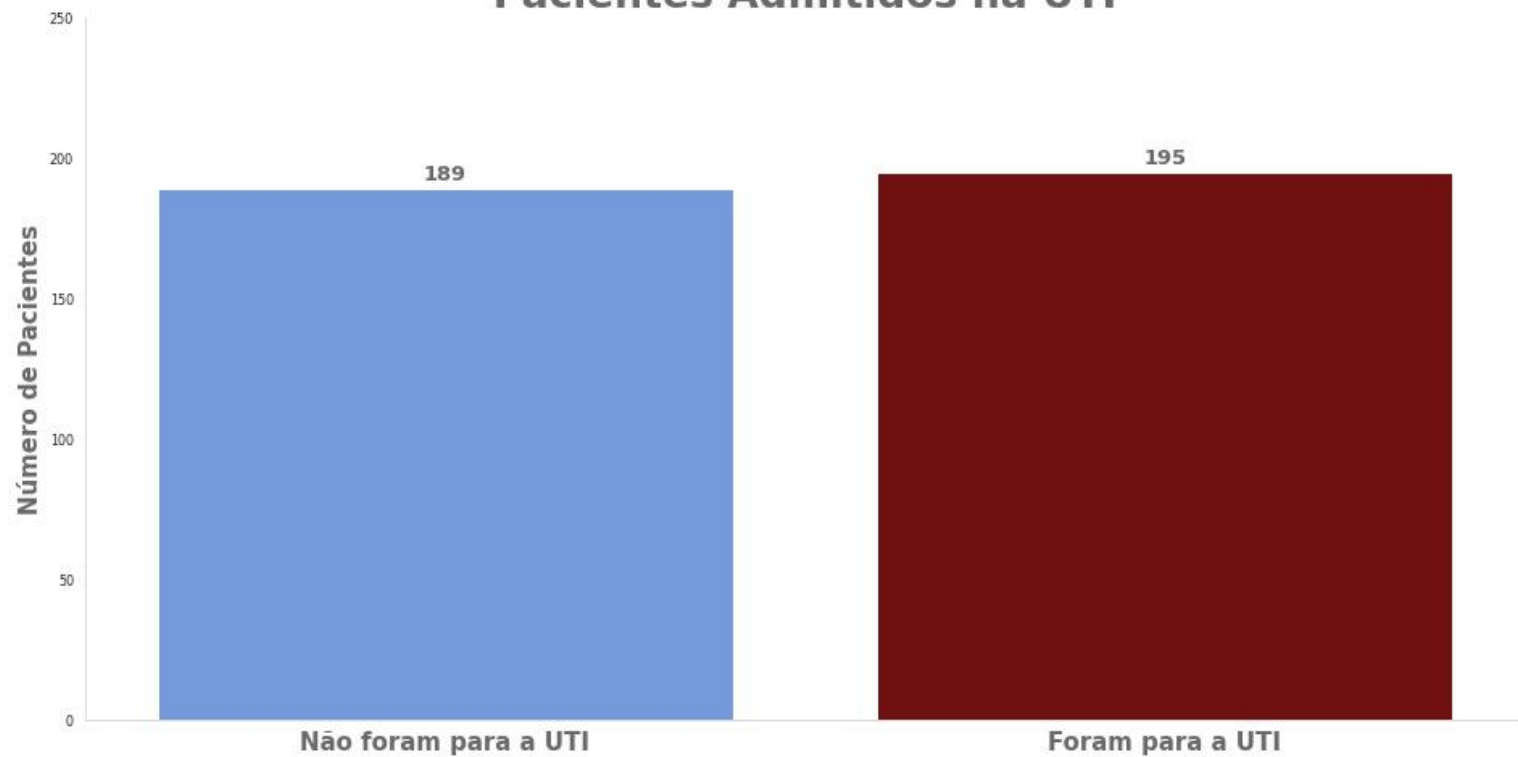
O projeto foi dividido em 3 partes:

1. **Tratamento dos dados;**
 - a. Preenchimento dos valores nulos;
 - b. Transformação dos dados em valores categóricos (onde necessário);
 - c. Selecionar a janela dos dados;
 - d. Remover dados altamente correlacionáveis;
2. **Análise Exploratória e Visualização dos Dados;**
3. **Modelos de Machine Learning;**



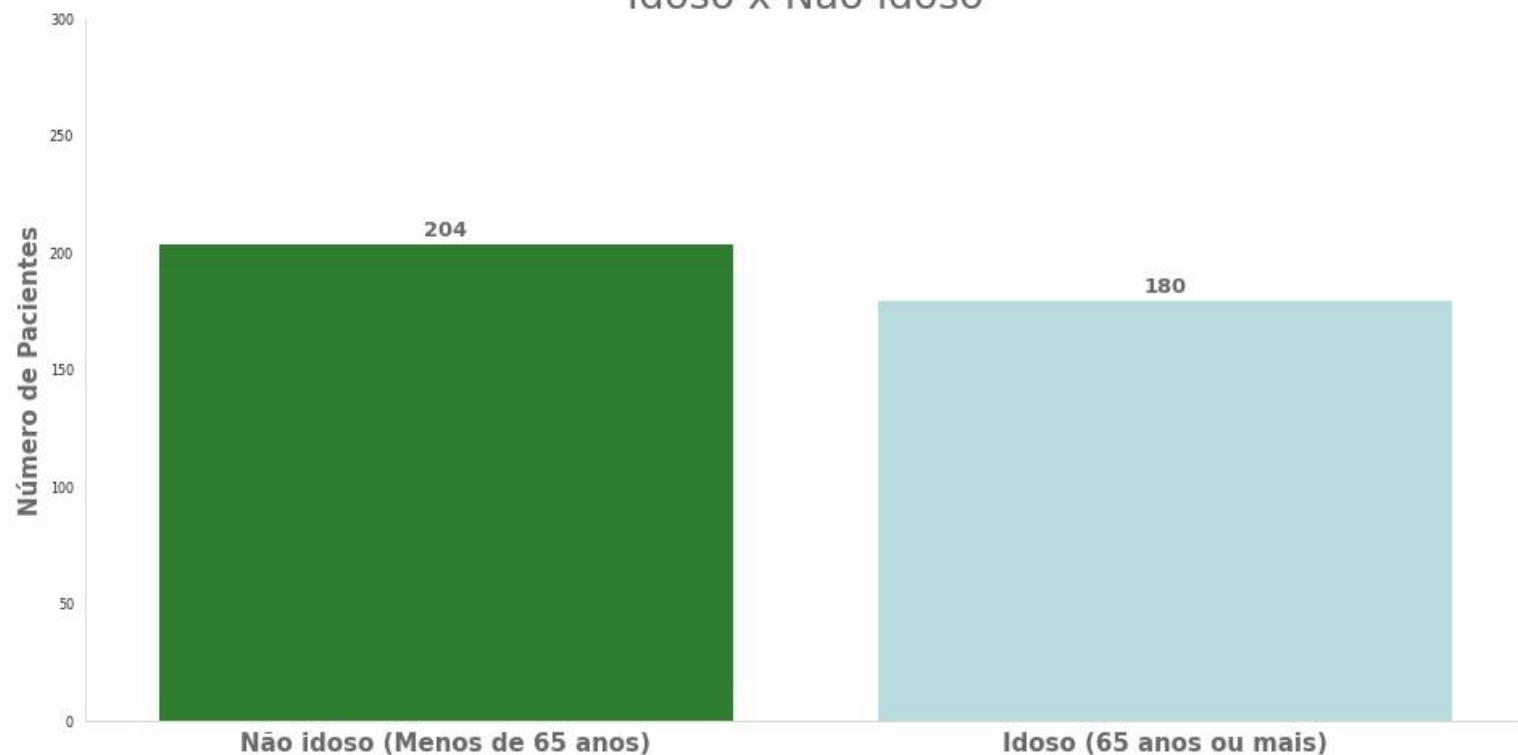
Informações dos Dados

Pacientes Admitidos na UTI



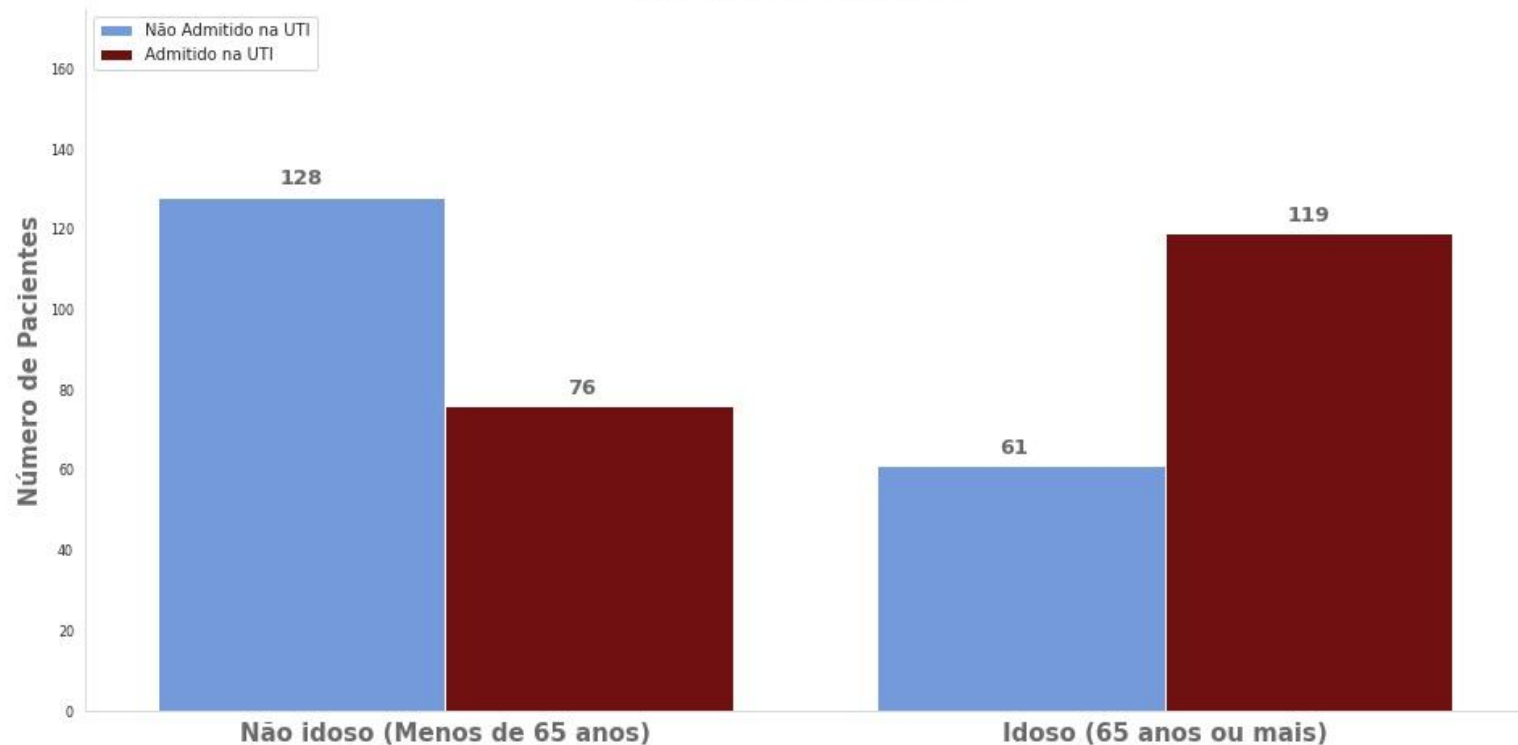
Idade dos Pacientes

Idoso x Não idoso



Admissão dos Pacientes na UTI

Idoso x Não idoso





Modelos de Machine Learning





Modelos Escolhidos

Os modelos de **Machine Learning** testados foram:

1. **Random Forest Classifier**
2. **Logistic Regression**
3. **SVC (Support Vector Classifier)**
4. **LinearSVC (Linear Support Vector Classifier)**



Métricas de Avaliação

Métricas de Avaliação analisadas:

- Curva ROC e AUC;
- Classification Report;
- Matriz de Confusão;



Melhores Modelos

Os modelos com os **melhores resultados** foram:

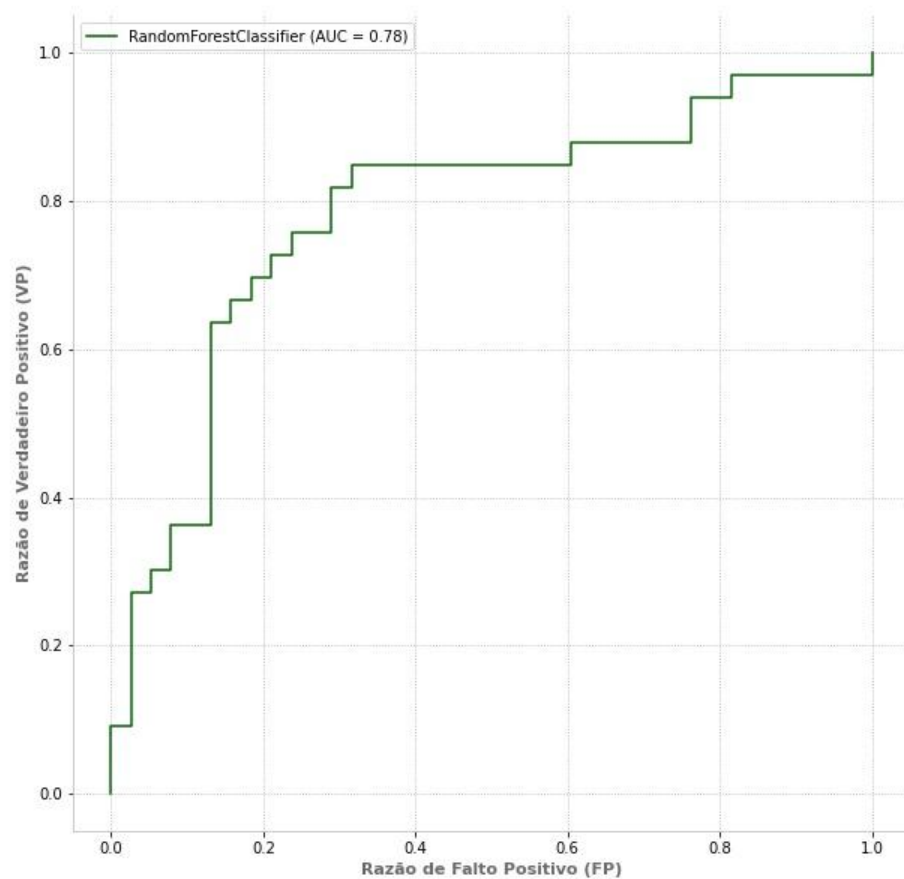
1. **Random Forest Classifier;**
2. **LinearSVC**



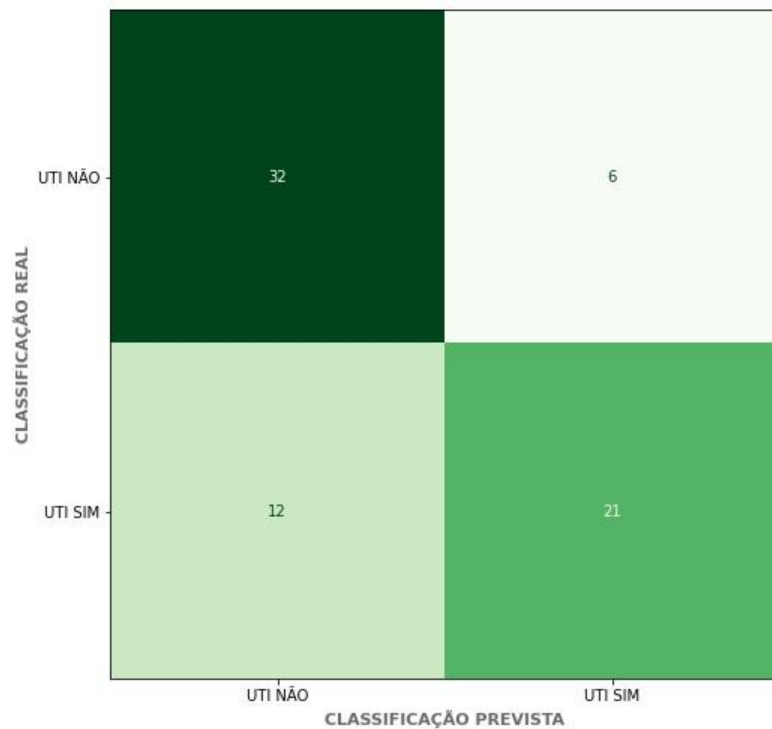
Métricas para o modelo Random Forest Classifier



Curva ROC para o modelo Random Forest



Matriz de Confusão para o modelo Random Forest

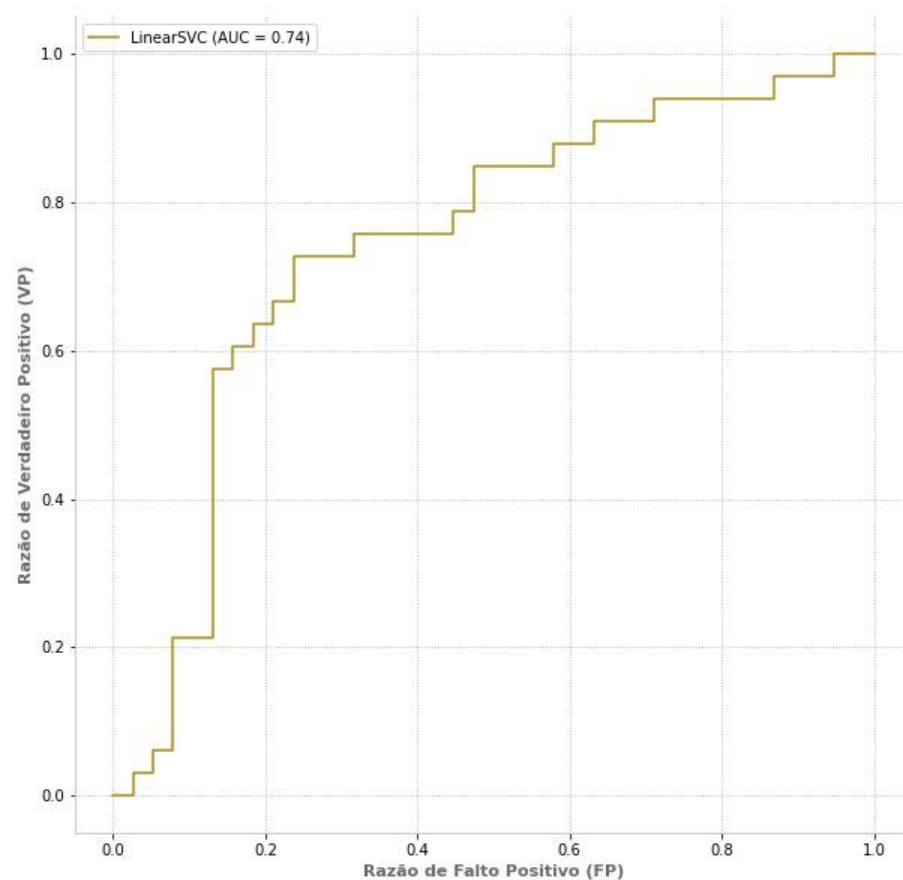




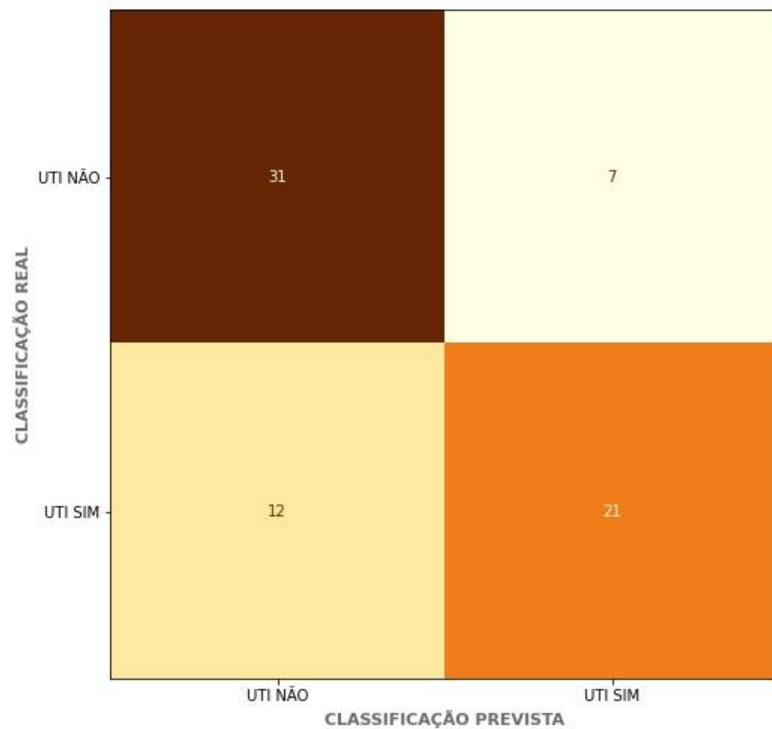
Métricas para o modelo LinearSVC



Curva ROC para o modelo LinearSVC



Matriz de Confusão para o modelo LinearSVC



Classification Report para os modelos testados

Métrica de Desempenho	Random Forest Classifier	LinearSVC	Logistic Regression	SVC
PRECISION(0)	0.73	0.72	0.72	0.69
PRECISION(1)	0.78	0.75	0.75	0.69
RECALL(0)	0.84	0.82	0.82	0.76
RECALL(1)	0.64	0.64	0.64	0.61
F1.Score(0)	0.78	0.77	0.77	0.72
F1.Score(1)	0.70	0.69	0.69	0.65
AUC TEST	0.80	0.77	0.77	0.77
AUC TREINO	0.98	0.85	0.85	0.86
AUC	0.78	0.74	0.74	0.75