Machine Learning Explainability with SHAP

O objetivo desse estudo é utilizar SHAP para fazer uma análise de quais variáveis tiveram maior influência na probabilidade de sobrevivência ao naufrágio (desafio Titanic).

Titanic: Machine Learning from Disaster

Historinha

Clássico! Desafio alguém que não chorou assitindo ao filme, mas de raiva! Aquele pedaço de madeira dava p/ os dois! kkk
Brincadeiras a parte...
O naufrágio aconteceu em 15 de abril de 1912, morreram 1502 pessoas de um total de 2224 passageiros. Alguns grupos de pessoas eram mais propensos a escaparem da morte do que outros. Por exemplo, mulheres, crianças e passageiros da 1ª Classe. Então, acho que da pra encontrar algum padrão que podemos extrair dos dados.

OBS

Lembrando que a ideia aqui é utilizar o método SHAP para explicabilidade do modelo!!! Por isso foco não é necessariamente ter a melhor acurácia de modelo, mas sim um valor suficiente para a garantia de que a explicabilidade possa ser confiável.

Dataset

- PassengerId: Número de identificação do passageiro
- Survived: flag marcando se foi sobrevivente ou não **--- 0 = No, 1 = Yes**
- Pclass: classe no navio **--- 1 = 1st, 2 = 2nd, 3 = 3rd**
- Name: nome do passageiro
- Sex: gênero
- Age: idade em anos
- SibSp: quantidade de irmãos / cônjuges a bordo do Titanic
- Parch: quanitdade de pais / filhos a bordo do Titanic
- Ticket: Número do ticket
- Fare: Tarifa do passageiro
- Cabin: Número da cabine
- Embarked: porto de embarcação **--- C = Cherbourg, Q = Queenstown, S = Southampton**

Notas:

sibsp: O conjunto de dados define as relações familiares desta forma ...

Irmão = irmão, irmã, meio-irmão, meia-irmã
Cônjuge = marido, esposa (amantes e noivos foram ignorados)

parch: O conjunto de dados define as relações familiares desta forma ...

Pai = mãe, pai
Criança = filha, filho, enteada, enteado
Algumas crianças viajavam apenas com a babá, portanto parch = 0 para elas.

Etapas

1. Qual o problema?
2. Carregando os dados
3. Análise Exploratória
4. Tratamento dos dados
5. Modelagem e Avaliação
6. **(e foco desse notebook) SHAP Explainability**

---

1. Qual o problema?

Desafio: O objetivo do desafio é utilizar os dados disponíveis para medir a probabilidade de sobrevivência dos passageiros do Titanic.

SHAP: O objetivo desse estudo é utilizar SHAP para fazer uma análise de quais variáveis tiveram maior influência na probabilidade de sobrevivência.

2.Carregando os Dados

Carregando base

```
train = pd.read_csv('titanic data/train.csv')
test = pd.read_csv('titanic data/test.csv')

print('train shape: ', train.shape)
print('test shape: ', test.shape)
```

```
train shape:  (891, 12)
test shape:  (418, 11)
```

3. Análise Exploratória

```
print('Shape:', train.shape)
train.head(2)
```

```
Shape: (891, 12)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	

Checando nulos

```
print('Dados faltantes:')
train.isnull().sum()
```

```
Dados faltantes:

PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

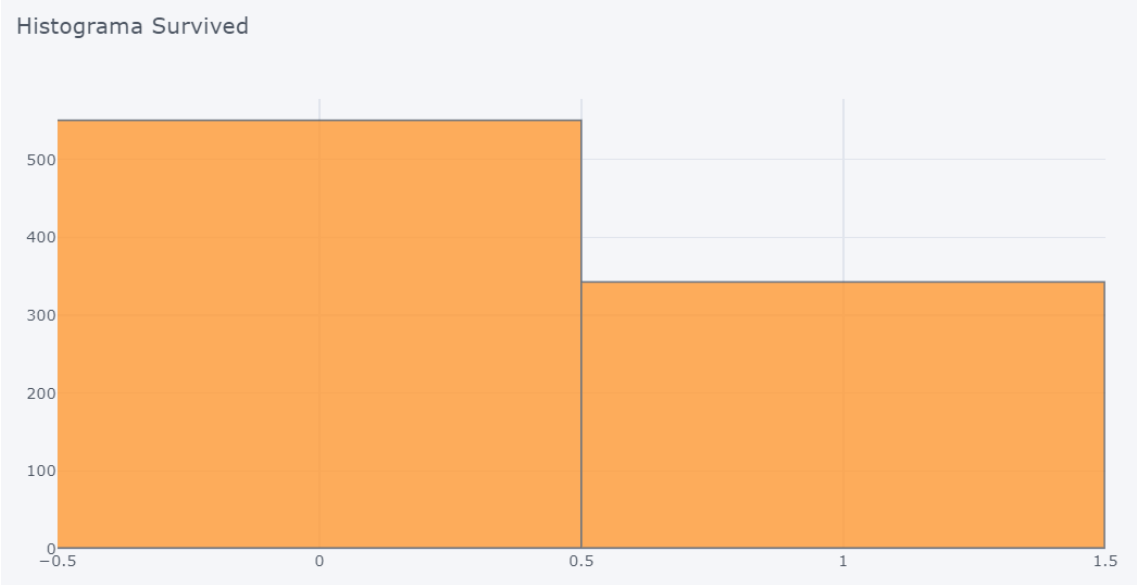
Temos 3 features com dados faltantes. Cabin (maior número), Age e Embarked.

Cabin: 77% faltante

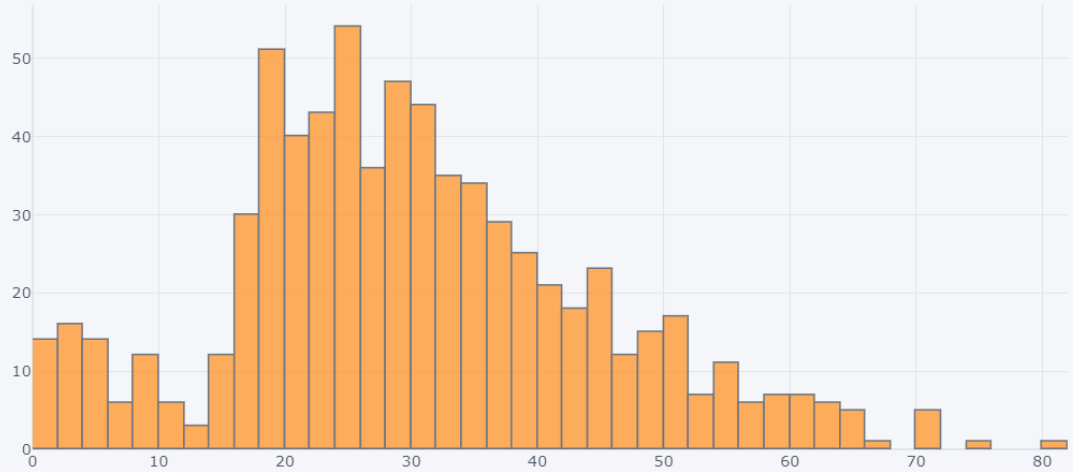
Age: 20% faltante

Embarked: 0.22%

HISTOGRAMAS



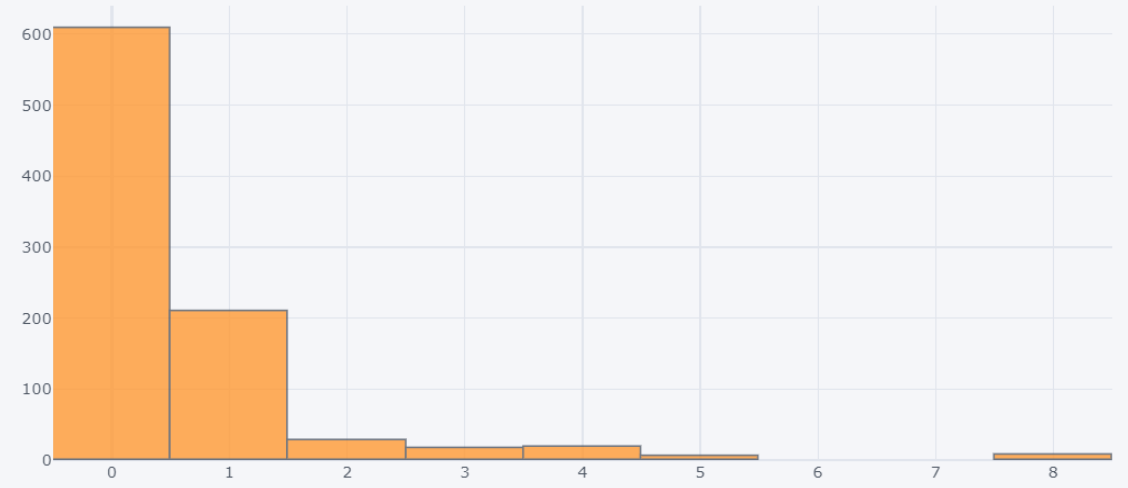
Histograma Age



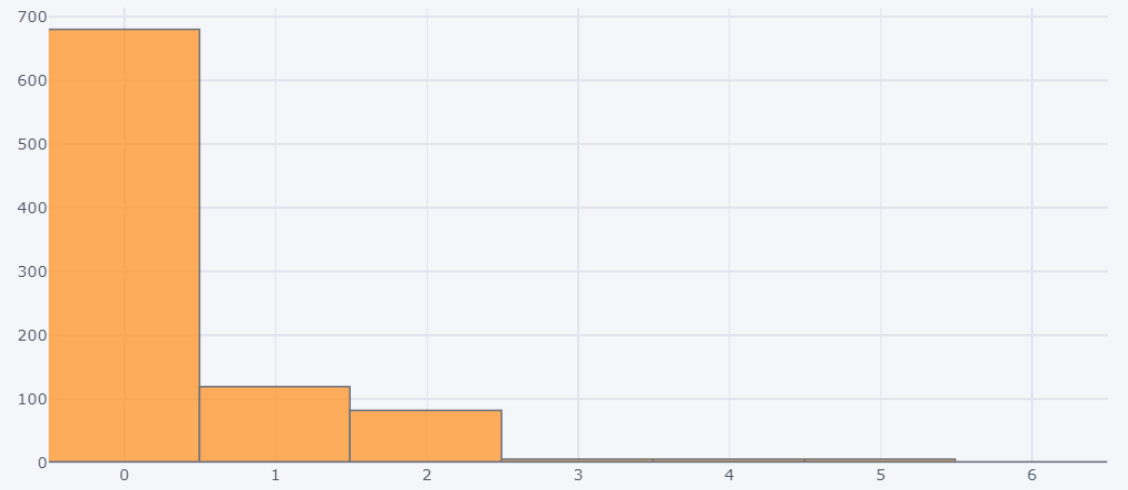
Histograma Pclass

Histograma Sex

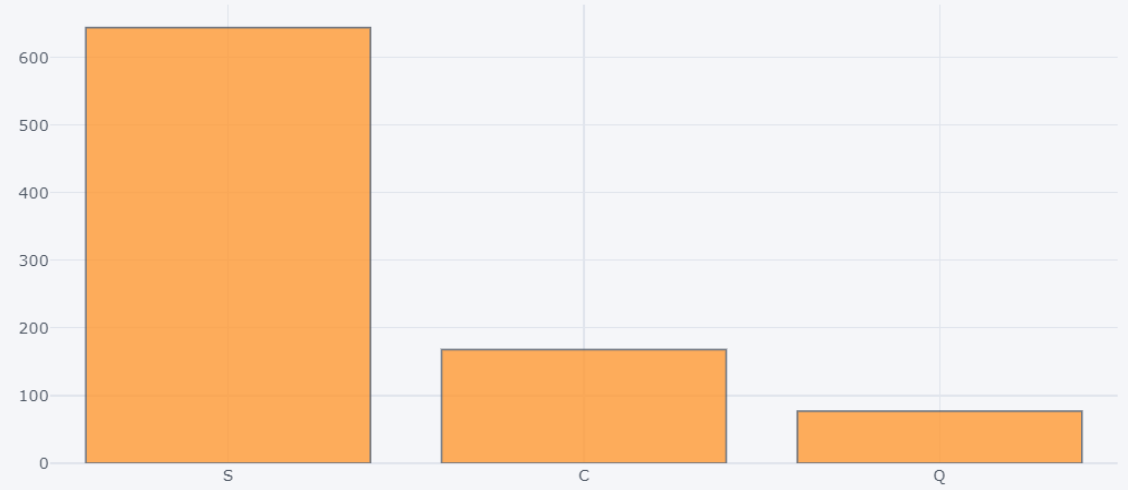
Histograma SibSp



Histograma Parch



Histograma Embarked



Comentários

Histograma Survived: Morreram mais pessoas do que sobreviveram.

```
62% Morreram, enquanto 38% sobreviveram (lembrando que aqui é só o dataframe de treino).
```

Histograma Age: vemos uma concentração de idade entre 18 e 38 anos

Histograma Pclass: mais pessoas na classe 3

Histograma Sex: mais homem (577 - 65% da base de treino)

Histograma SibSp: maior parte da base (68%) não tem irmãos/cônjuges a bordo. 23% tem apenas 1.

Histograma Parch: maior parte da base (76%) não tem pais/filhos a bordo. 13% tem apenas 1.

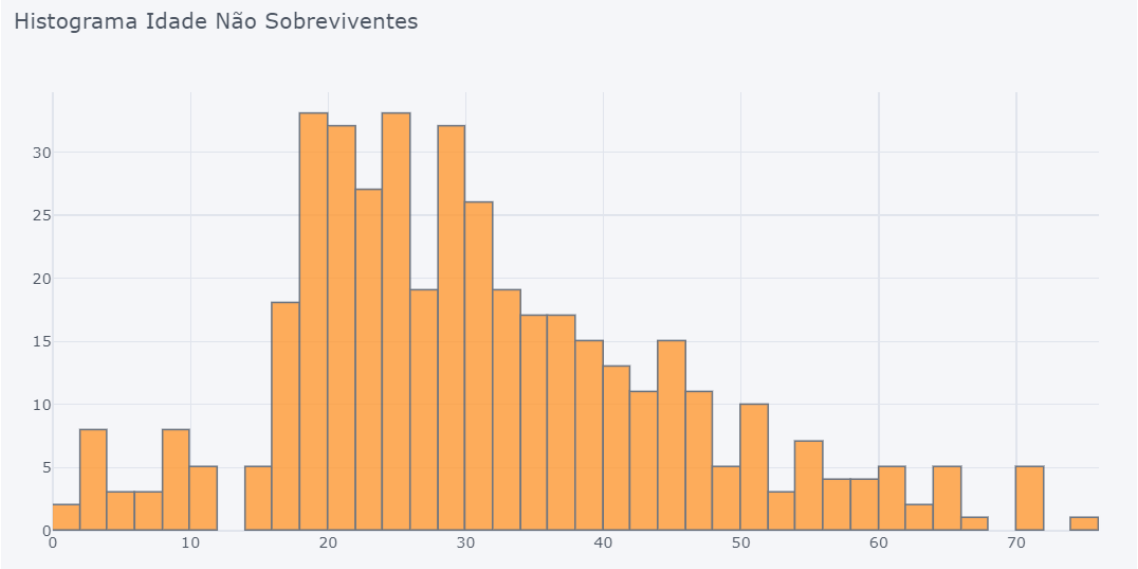
Histograma Embarked: maior parte dos passageiros (72%) embarcaram Southampton

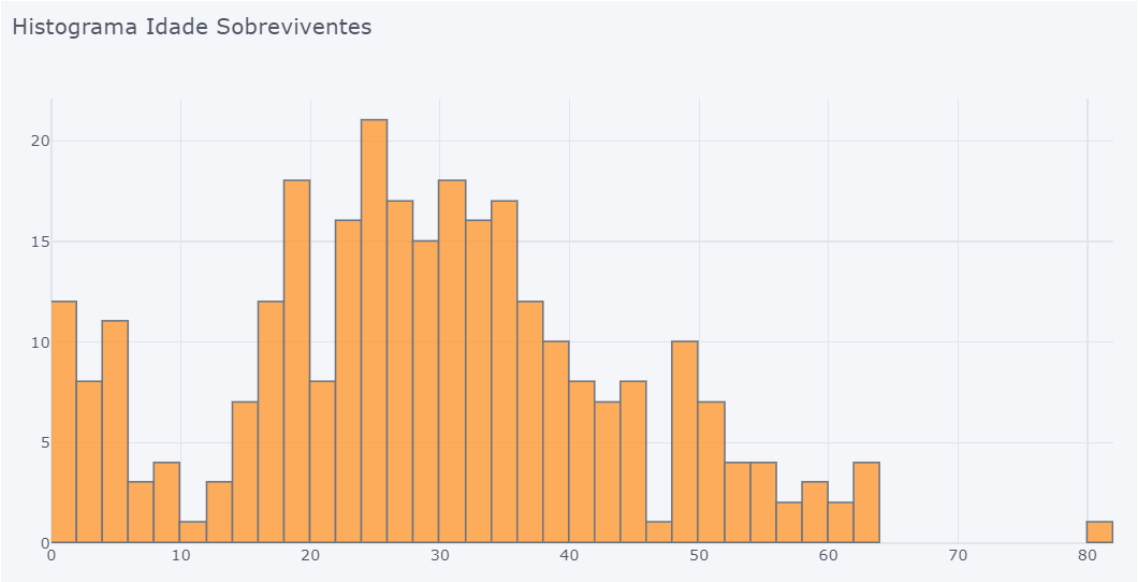
---

Visualizando histogramas considerando se sobreviveu ou não, para entendermos se classe, sexo, etc..podem ter influenciado na probabilidade de sobreviver ou não

---

IDADE

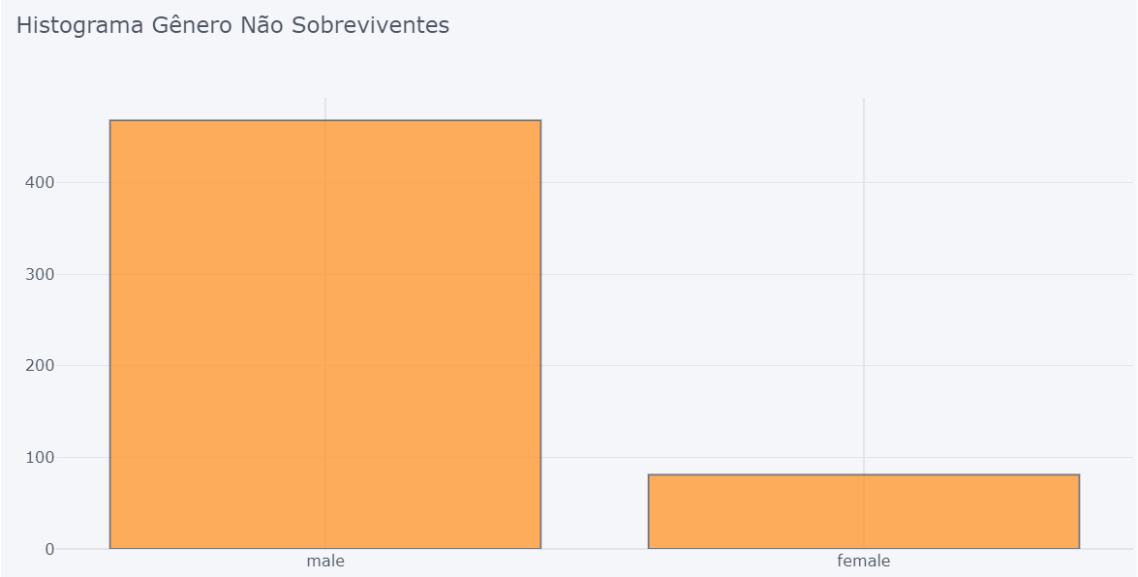


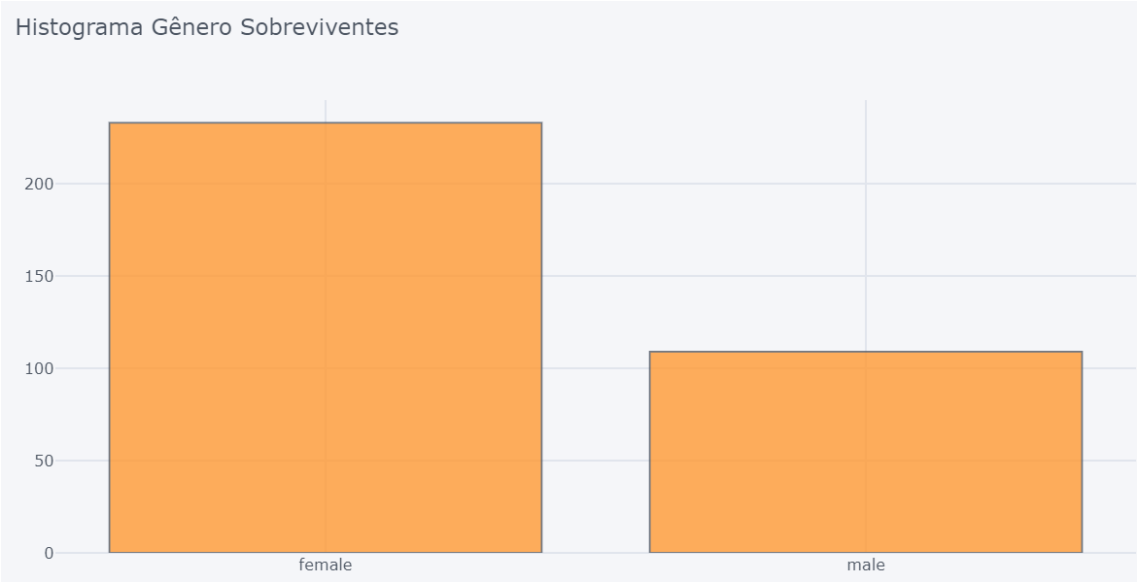


Comentário

Analisando a idade, é possível obervar que no histograma de so
breviventes, temos um "pico" no início do gráfico. O que indic
a uma hipótese de que crianças teriam mais chance de sobrevive
ncia.

GÊNERO



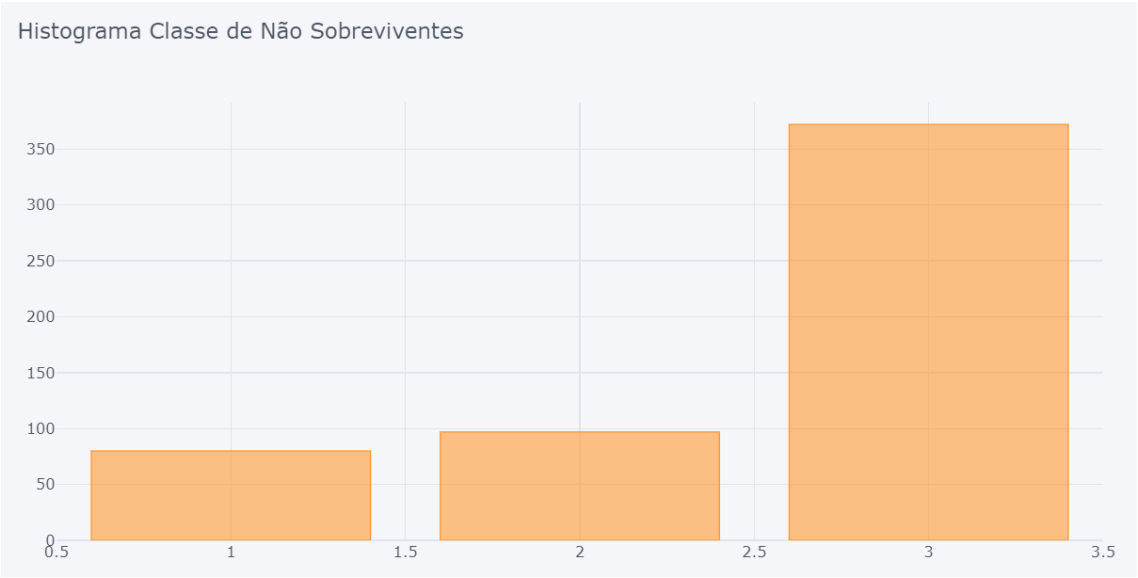


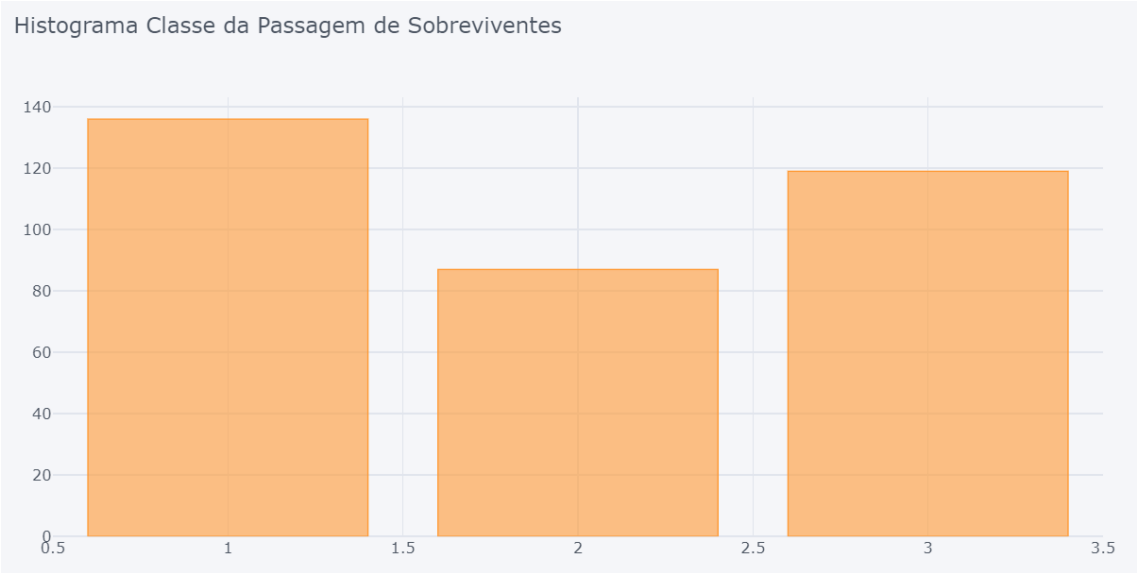
Comentário

É possível observar que os gráficos se "invertem". Quando olhamos para os passageiros que NÃO sobreviveram, a maioria é homem. Já quando olhamos para os passageiros do grupo que sobreviveram, temos mais mulheres. 233 mulheres sobreviveram, e apenas 109 homens. O que indicada que a hipótese de que as mulheres tinham mais chance de sobrevivência está correta.

Até aqui se confirma a frase tão falada no filme "Crianças e Mulheres primeiro" (para os botes salva vidas)

CLASSE





Comentário

No grupo de NÃO sobreviventes se destaca a classe 3. Já no grupo de sobreviventes observamos que o classe 1 predomina. Hipotese já imaginada, pobre tem mais chance de morrer que rico :(

4. Tratamento dos dados

Tratando nulos das colunas Cabin (maior número), Age e Embarked. (já vistos anteriormente no código)

Age: 20% faltante
Embarked: 0.22%

Idade vou colocar o valor da mediana.
Embarque vou colocar o valor com maior frequência, que é S

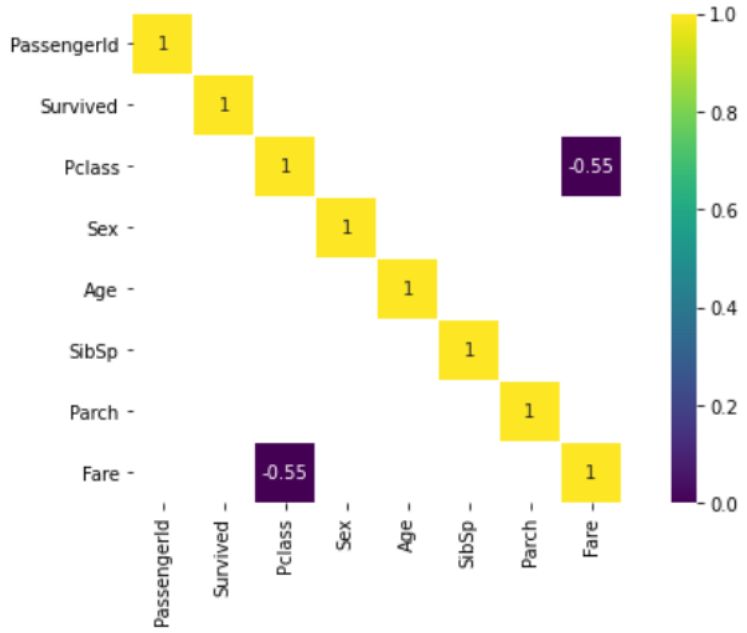
Na base de teste tem um da faltante na coluna Fare, vou substituir pela média

Convertendo Sex em 0 e 1

Convertendo Embarked em 0, 1 e 2

Float to int

Correlação



5. Modelagem e Avaliação

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_ma
```

Desconsiderando algumas variáveis

Obs: Cabin poderia ter uma análise mais detalhada, as letra s tem uma ordem dentro do navio. Mas vou deixar pra um outr o momento. Lembrando que o intuito aqui é o entendimento do SHAP.

Separando em treino e teste

```
X_train = df_train.drop(['PassengerId', 'Name', 'Ticket', 'Cabin',
y_train = df_train['Survived']
X_test  = df_test.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'],
```

Fit e score

```
model = RandomForestClassifier(random_state=0)

# Estimando o modelo com a base de treino
model.fit(X_train, y_train)

# verificar a acurácia do modelo
acuracia = round(model.score(X_train, y_train) * 100, 2)
print('Acurácia com RandomForestClassifier:', acuracia, '%')
```

```
Acurácia com RandomForestClassifier: 96.52 %
```

Vou utilizar o K-fold para uma classificação mais confiável e realista

K-Fold Cross Validation

K-Fold Cross Validation divide aleatoriamente os dados de trei
namento em K subsets chamados folds. Se o k = 3, por exemplo,
nosso modelo Random Forest seria treinado e avaliado 5 vezes.
Por exemplo, em uma etapa o modelo seria treinado com o subcon
junto 1 e 2 e avaliado com o subconjunto 3. Em uma segunda eta
pa treinado com o subconjunto 1 e 3, e avaliado com o 2. E uma
terceira etapa treinado com o 2 e 3 e avaliado com o 1.
O resultado neste caso seria uma matriz com 3 "notas" diferent
es. Aí podemos calcular a média e o desvio padrão delas, por e
xemplo.

Teste com K = 10

```
1  from sklearn.model_selection import cross_val_score
2
3  model = RandomForestClassifier(n_estimators=100)
4
5  scores = cross_val_score(model, X_train, y_train, cv=10, scoring =
6
7  print('Score:', scores)
8  print('\nMédia:', round(scores.mean(),2) * 100 , '%')
9  print('\nDesvio Padrão:', round(scores.std(),2) * 100 , '%')
```

```
Score: [0.73333333 0.79775281 0.7752809  0.82022472 0.88764045 0.8202
2472
 0.86516854 0.76404494 0.82022472 0.84269663]

Média: 81.0 %

Desvio Padrão: 4.0 %


Precision: 0.7439759036144579
Recall: 0.7222222222222222
F1: 0.7329376854599406

array([[464,  85],
       [ 95, 247]], dtype=int64)
```

Bom, a partir daqui poderia fazer ajuste dos hiperparâmetros buscando melhorar a performance do modelo. MAS, como mencionado várias vezes, a ideia aqui é utilizar o método SHAP. Como a acurácia obtida já está 'boa', podemos confiar na explicabilidade das variáveis. Então, a partir daqui não me importarei mais com as métricas e partirei para a análise com SHAP

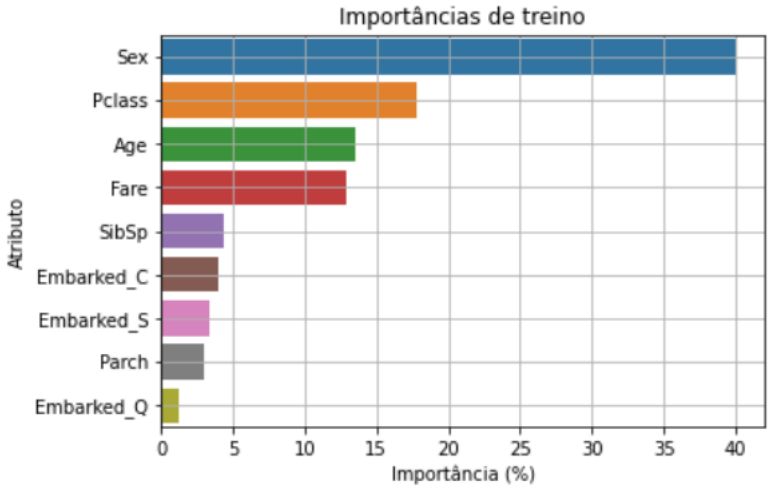
```
model.fit(X_train, y_train)
pred = model.predict(X_test)
```

```
submission = pd.DataFrame({
        'PassengerId': df_test['PassengerId'],
        'Survived': pred
    })
submission.to_csv('submission.csv', index=False)
```

Essa submissão gerou um score de 75% no kaggle

---

6. SHAP Explainability



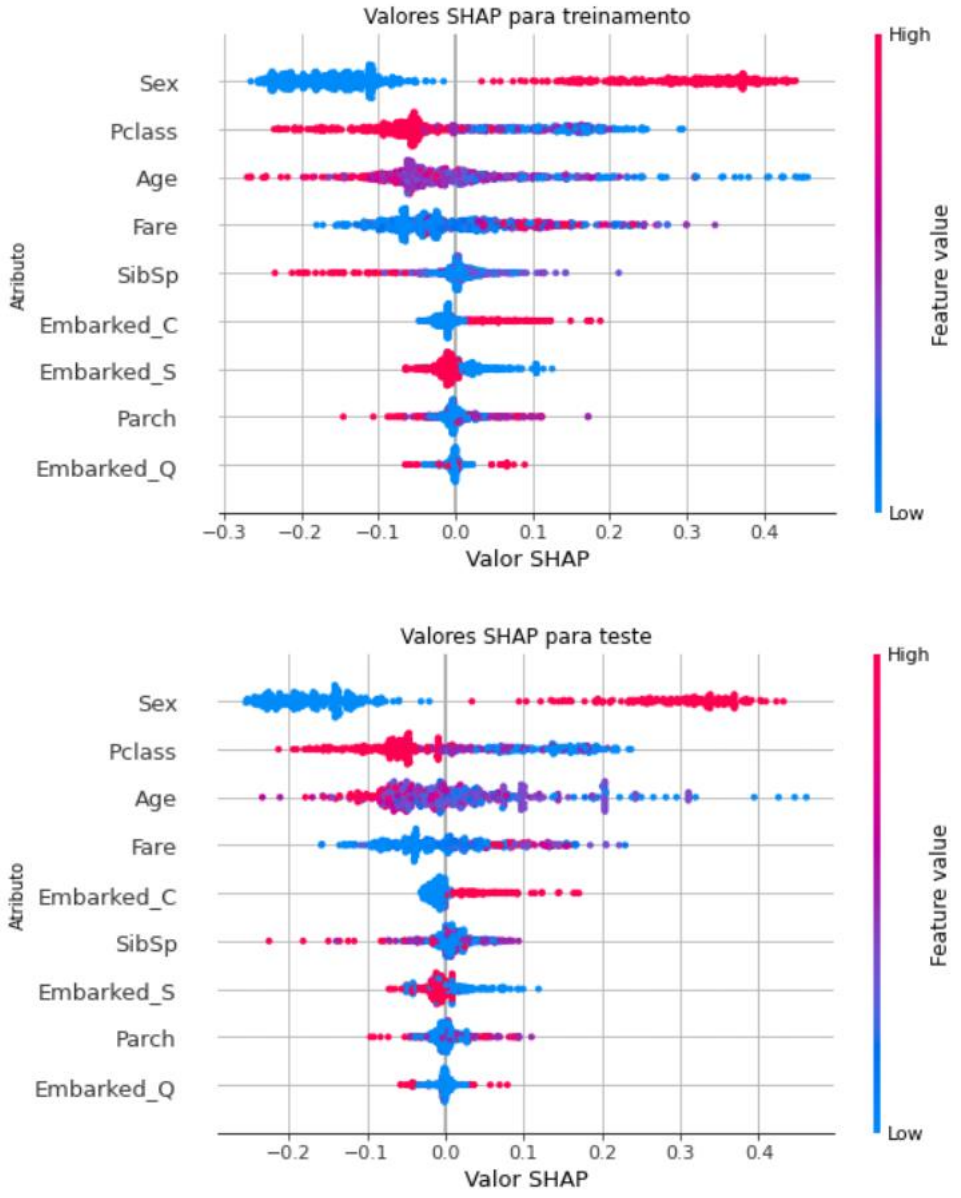
Comentários da visualização acima

fica claro a importância da variável 'Sex' no modelo. Segu
ido da variável classe e da idade. O que nos lembra das hipóte
ses levantadas.

Mulheres e crianças sobreviveram mais?

Classes mais baixas morreram mais?

SUMMARY PLOT



Comentários da visualização acima

Como ler esse tipo de gráfico: Cada pontinho dess é uma amostra classificada. Eixo 'x' dessa visualização é o Valor SHAP, que está relacionado com o valor de saída do modelo. Então quanto mais pra direita do eixo x (valor SHAP), maior a chance do passageiro ser classificado como sobrevivente (1), quanto mais pra esquerda, maior a chance do passageiro ser classificado como NÃO sobrevivente (0). Lembrando: 0: não sobrevivendo **---** 1:sobrevivente

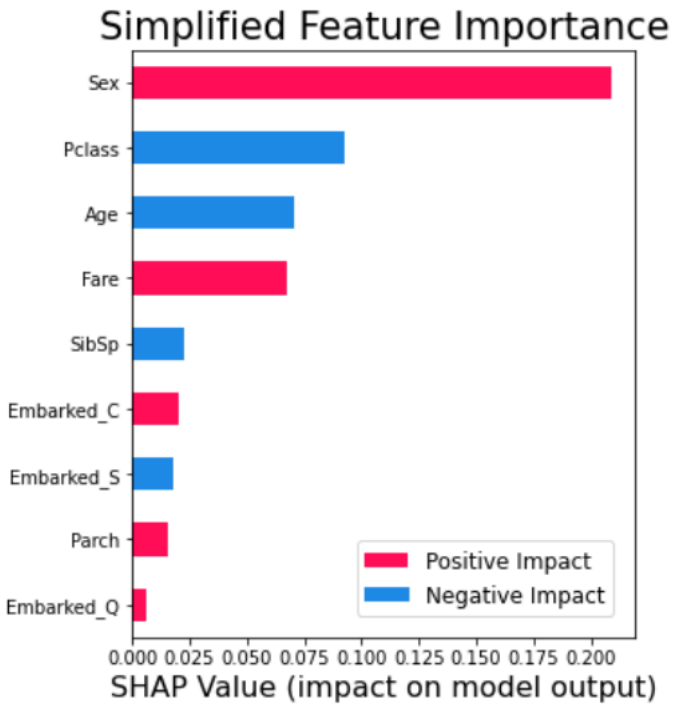
```
    SEX
    Lembra que '0' é homem e '1' é mulher?
    Então, podemos ver que quanto maior o valor da feature, ma
is vermelho é. Pensando na variável 'Sex', se for vermelho é 1
(mulher), azul é 0 (homem)
    Fica clara essa separação nessa visualização. Quanto mais
vermelho (mulher), mais o gráfico tende para a direita do eixo
'x' (valor SHAP). Ou seja, quanto mais vermelho (mulher), mais
chances de sobreviver (direita do eixo x).
```

```
    CLASSE
    Lembra que comentei a hipotese de que pobre morreu mais?
    Então, pensando na variável 'Pclass', quanto mais vermelh
o, maior a classe (mais pobre kkk). E quanto mais azul, menor
a classe (first class, mais rico).
    A leitura é: Quanto mais vermelho, mais as predições estão
pro lado esquerdo do gráfico (mais chances de não sobreviver).
Ou seja quanto mais vermelho, maior a classe (mais pobre), mai
s pro lado esquerdo do gráfico, menos chances de sobreviver (s
er classificado como Survived = 0)

    IDADE
    Lembra que comentei da hipotese que crianças mais novas te
riam mais chances de sobrevivier?
    Então, podemos ver que quanto menor a idade (mais azul), m
ais pro lado direito do gráfico está (mais chances de ser clas
sificado como sobrevivente)

    FARE
    Caracterítica que anda junto com classe no navio. Quanto m
ais caro o bilhete, melhor a classe.
    E a gente consegue ver que as bolinhas azuis pra variável
'Fare' (valor de bilhete mais barato) estão do lado esquerdo d
o gráfico, ou seja, tem mais probabilidade de serem classifica
dos como não sobreviventes.
```

Simplified plot



Comentários da visualização acima

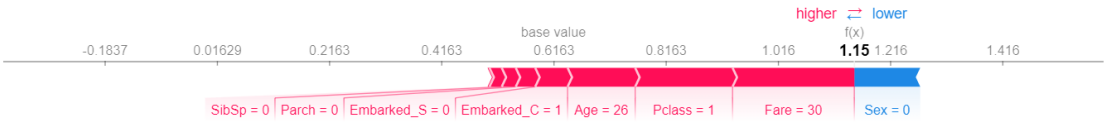
Como ler esse tipo de gráfico: Visualização que resume a anterior. Vermelho tem impacto positivo, ou seja, quanto maior, mais influencia para a classificação ser 1 (sobrevivente). Azul tem impacto negativo, ou seja, inverso, quanto maior, mais influencia para a classificação ser 0 (não sobrevivente)

```
Ex:
SEX ta como vermelho, então de uma forma geral, quanto maior a
variável 'Sex' (mulher), maior a chance de sobreviver (classe
1)
PCLASS ta como azul, então de uma forma geral, quanto menor a
variávle 'Pclass' (mais perto da primeira classe, público ric
o), maior a chance de sobreviver (classe 1)
```

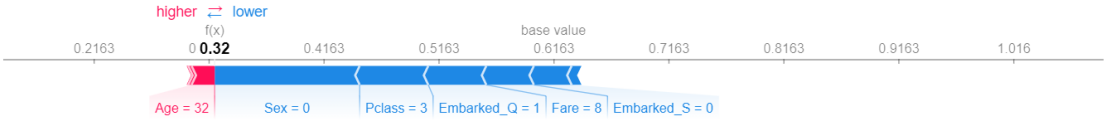
Passageiros específicos

Observando com outro tipo de visualização. Agora pegando casos isolados

Line 889



Line 890



Comentário da visualização acima

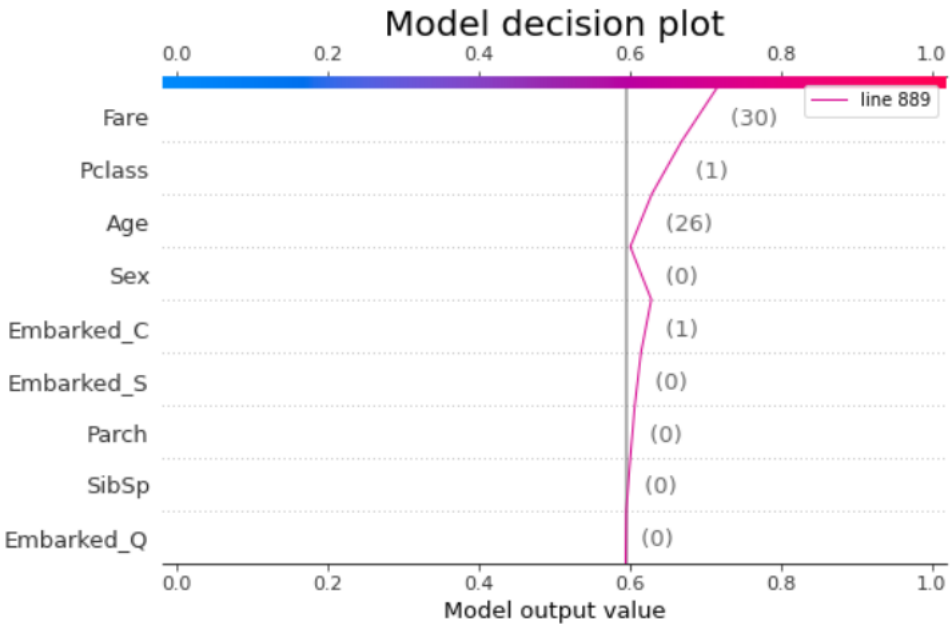
Como ler esse tipo de gráfico: Nesse tipo de gráfico podemos analisar casos individualmente, para uma amostra (passageiro) específico. Tamanho da barrinha de cada variável define o impacto dela no modelo, quanto maior, mais impacto. A cor define se impactou positivamente (vermelho) "empurrando pra direita" ou negativamente (azul) "empurrando pra esquerda". No gráfico, quanto mais o valor for "empurrado" para a direita, mais probabilidade de ser classificado como 1 (sobrevivente), quanto mais o valor for "puxado" pra esquerda, maior a probailidade de ser classificado como 0 (não sobrevivente).

Análise para o passageiro da linha 889 Podemos ver que as informações que mais tiveram peso para a classificação foram "Fare", seguido "Pclass" e "Age", todas na cor vermelha, ou seja, essas variáveis influenciaram na probabilidade do passageiro ser classificado como 1 (sobrevivente). E a única variável que "puxou" o valor para esquerda foi 'Sex', o fato de o passageiro ser homem. Neste caso, "apesar" dele ser homem, foi classificado como sobrevivente, porque o fato de ser novo (26 anos), ter pago 30 de taria para estar na classe 1 (first class :)) fez ele ser "sortudo" e ser classificado como sobrevivente. Rico tem sorte mesmo, né?

Análise para o passageiro da linha 890 Neste exemplo temos uma classificação 0 (não sobrevivente). A variável de maior peso foi "Sex", seguido de 'Embarked_Q" e "Pclass". O fato de ser homem, tem embarcado no porto "Q" e estar na classe 3, contribuiu para a classificação como 0 (não sobrevivente). A variável que "puxou para direita (para ser classificado como sobrevivente) foi "Age", mas apesar dele ter 32 anos, os outros fatores impactaram mais.

Decision Plot

Observando com outro tipo de visualização. Ainda pegando casos isolados, aqui continuando com o passageiro da linhas 889

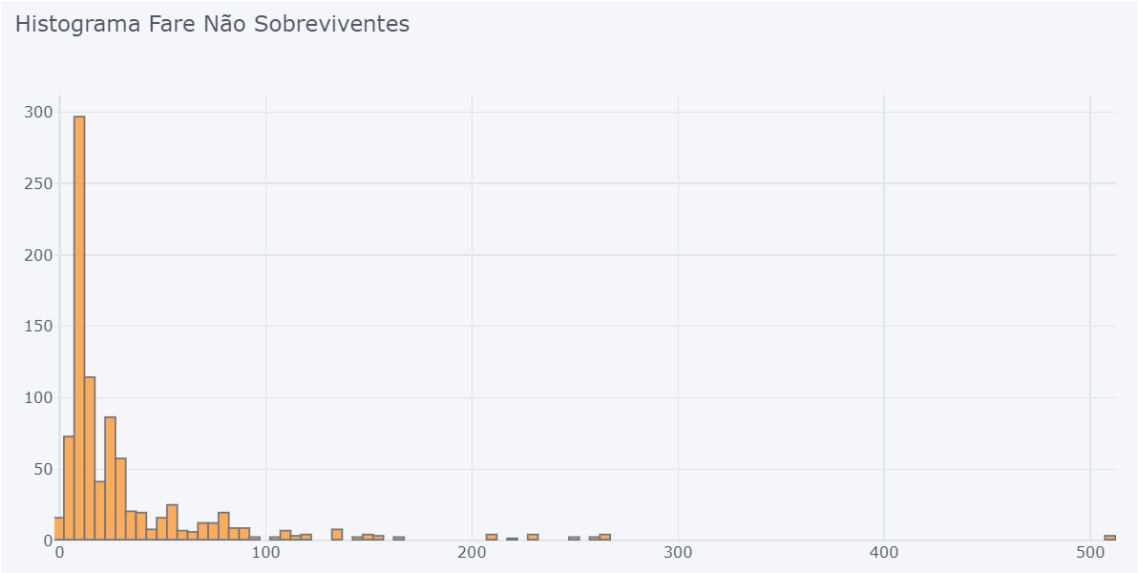


Comentário da visualização acima

Como ler esse tipo de gráfico: Nesse tipo de gráfico podemos analisar casos individualmente, para uma amostra (passageiro) específico. Nele conseguimos ver o impacto de cada variável na classificação, a reta vai sendo "puxada" para a direito ou para a esquerda, conforme o valor da variável, até chegar na classificação. Quanto mais o valor for "puxado" para a direita, mais probabilidade de ser classificado como 1 (sobrevivente), quanto mais o valor for "puxado" pra esquerda, maior a probailidade de ser classificado como 0 (não sobrevivente).

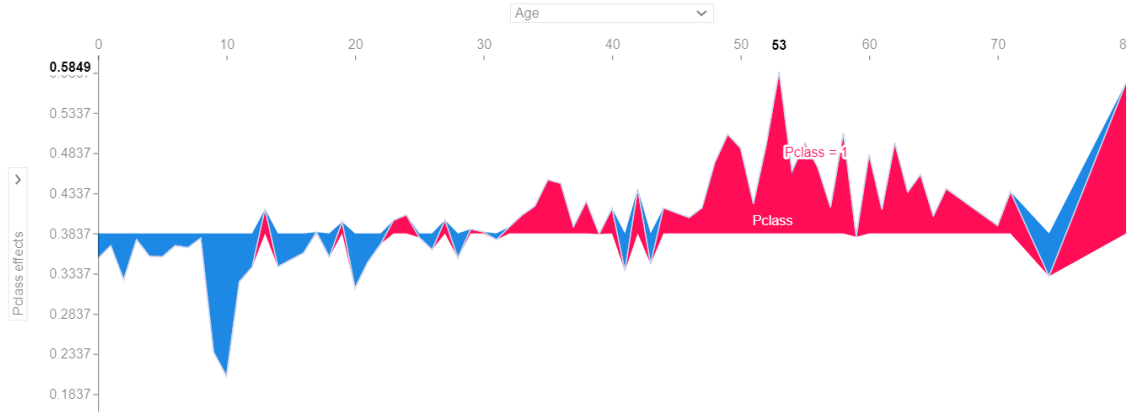
Análise para o passageiro da linha 889 Podemos ver que as informações da parte inferior da visualização, "Embarked_Q", "Embarked_S", "SibSp" e "Parch" não influenciaram muito, a reta permaneceu perto do eixo central. Quando seguimos a reta e chegamos nas variáveis "Embarked_C" = 1 e "Age" = 26, podemos observar uma leve "puxada" para o lado direito do gráfico (probabilidade de ser classificado como sobrevivente), mas seguindo para a próxima variável, "sex" = 0 (homem), faz a reta ser puxada para a esquerda da visualização (probabilidade de ser classificado como não sobrevivente). Após isso, "Pclass" = 1 (first class) e "Fare" = 30 (valor maior que a maioria dos passageiros pagaram (veja a distriuição no plot abaixo), puxaram a reta mais ainda para a direita, o e o passageiro teve classificação 1 (sobrevivente).

Distribuição da variável "Fare"



Essa é mais uma visualização legal do SHAP

É uma visualização iterativa, podemos escolher qual variável vai estar em cada eixo e analisarmos a correlação entre ela s. f(x) é a probabilidade.



Dependense plot

Aqui cruzamos as variáveis entre elas, p/ tentar ver como elas se correlacionam entre si.

Exemplos: