

1.3 Titanic

1. Faça uma análise inicial das variáveis (exploratory data analysis) apresentadas em:
<https://www.kaggle.com/c/titanic/download/train.csv>

Resposta: Imagens na pasta “imagens”, geradas no código analise.py

2. Tente prever a variável Survival, quais passageiros estavam mais propensos à sobrevivência, baseado nos atributos estudados na questão anterior.

Resposta: Com base na análise feita nos atributos estudados na questão anterior, os passageiros mais propensos à sobrevivência são aqueles que o atributo “Pclass” é igual à 1, ou seja, que estavam na primeira classe; os que tinham idade (atributo “Age”) entre 0 e 4 anos, os que pagaram maior valor de passagem (atributo “Fare”). Além disso, é importante notar que a maior partes dos sobreviventes são do sexo feminino, até em percentual, onde aproximadamente 74.20% das mulheres que embarcaram sobreviveram, enquanto apenas 18.89% dos homens saíram com vida.

Obs: Algoritmos que ajudaram também na interpretação desta questão, estão no código previsao.py. A previsão realizada pelo melhor modelo treinado se encontra no arquivo previsao.csv e foi analisada no código teste.py, no qual o modelo foi usado para prever a variável “Survived” dos dados encontrados em test.csv (que por sua vez se encontra na pasta “bases”)