**Visão geral**

Os dados foram divididos em dois grupos:

conjunto de treinamento (train.csv)

conjunto de teste (test.csv)

O conjunto de treinamento deve ser usado para construir seus modelos de aprendizado de máquina. Para o conjunto de treinamento, fornecemos o resultado (também conhecido como “verdade básica”) para cada passageiro. Seu modelo será baseado em “características” como gênero e classe dos passageiros. Você também pode usar a engenharia de recursos para criar novos recursos.

O conjunto de testes deve ser usado para ver o desempenho do seu modelo em dados não vistos. Para o conjunto de testes, não fornecemos informações básicas para cada passageiro. É seu trabalho prever esses resultados. Para cada passageiro no conjunto de teste, use o modelo que você treinou para prever se eles sobreviveram ou não ao naufrágio do Titanic.

**Data Dictionary**

|  |  |  |  |
| --- | --- | --- | --- |
| **VARIÁVEL** | **TRADUÇÃO** | **DEFINIÇÃO** | **CHAVE** |
| PassengerId | ID Passageiro |  |  |
| Survived | Sobreviveu |  | 0= Não 1 = Sim |
| Pclass | Classe |  | 1°, 2º e 3º |
| Name | Nome |  |  |
| Sex | Sexo |  |  |
| Age | Idade |  |  |
| SibSp |  | Irmãos/cônjuges a bordo do Titanic |  |
| Parch |  | Pais/filhos a bordo do Titanic |  |
| Ticket |  | Número Bilhete |  |
| Fare | Tarifa | Tarifa Passageiro |  |
| Cabin | Cabine | Número da Cabine |  |
| Embarked | Embarcado | Porto de Embarque | C = Cherbourg  Q = Queenstown  S = Southampton |

Notas Variáveis

Classes

1st = Alta  
2nd = Média

3rd = Baixa

Idade

A idade é fracionária se for menor que 1. Se a idade for estimada, está na forma de xx,5

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 babá, portanto parch=0 para elas.

**Formato do arquivo de envio**

Você deve enviar um arquivo csv com exatamente 418 entradas mais uma linha de cabeçalho. Seu envio mostrará um erro se você tiver colunas ou linhas extras (além de PassengerId e Survived).

O arquivo deve ter exatamente 2 colunas:

PassengerId (classificado em qualquer ordem)

Sobreviveu (contém suas previsões binárias: 1 para sobreviventes, 0 para falecidos)