

Assignment 1 – Python and Machine Learning Basics

O objetivo é prever se um passageiro sobreviveu ou não no desafio do Titanic do Kaggle, com base em atributos como idade, sexo, classe do passageiro, onde embarcou e assim por diante.

Primeiro, faça login no Kaggle e vá para o [desafio do Titanic](#) para baixar train.csv e test.csv. Os dados já estão divididos em um conjunto de treinamento e um conjunto de teste. No entanto, os dados de teste não contêm os rótulos: seu objetivo é treinar o melhor modelo possível usando os dados de treinamento, então faça suas previsões sobre os dados de teste e carregue-os no Kaggle para ver sua pontuação final.

Primeiro, avalie os atributos do conjunto de treinamento. Alguns atributos contêm dados faltantes. Isso indica que pode ser desnecessário incluí-los no modelo.

Em seguida, avalie o desempenho de alguns modelos vistos em aula.

Para melhorar esse resultado, você pode:

- Compare mais modelos e ajuste hiperparâmetros usando validação cruzada e grid Search;
- Faça pré-processamento das features, por exemplo:
 - substituir SibSp e Parch por sua soma,
 - tente identificar partes de nomes que se correlacionam bem com o atributo Sobrevivido (por exemplo, se o nome contém "Condessa", então a sobrevivência parece mais provável),
- Tente converter atributos numéricos em atributos categóricos: por exemplo, diferentes grupos de idade tinham taxas de sobrevivência muito diferentes. Então pode ajudar a criar uma categoria de faixa etária e usá-la em vez da idade. Da mesma forma, pode ser útil ter uma categoria especial para pessoas que viajam sozinhas, uma vez que apenas 30% delas sobreviveram.