

integrantes: Antônio Humberto

competição: Titanic - Machine Learning from disaster

objetivo do desafio: desenvolver um modelo capaz de prever o destino dos passageiros do naufrágio do titanic utilizando as informações passadas nos datasets disponibilizados

link competição: <https://www.kaggle.com/competitions/titanic/overview>

Análise exploratória de dados: como se tratava de um acontecimento bem conhecido e documentado a maioria das suposições vieram de conhecimento comum como: mulheres e crianças teriam mais chance de sobreviver por terem prioridade nos botes de resgate, a posição da cabine do passageiro no navio, passageiros ricos e status social foram as suposições iniciais. O fato dos datasets serem bem limpos e vastamente trabalhados facilitou na hora de pesquisar por referências em outros trabalhos.

Explicação do modelo: Para o modelo foi utilizado Random Forest já que facilita a entender as features, funciona bem com poucos dados e consegue aprender relações complexas sozinho. Também foi usado Cross-Validation para validação local. Feature Engineering principalmente para agrupar algumas variáveis mais dispersas em uma única mais impactante para o modelo ou gerar uma nova coluna baseado nos dados do dataset.

Implementação: no total foram feitas 6-7 versões para o modelo, com a primeira seguindo o tutorial disponível no próprio kaggle, aos poucos foram adicionando mais variáveis para o modelo considerar e melhorar o score. Por volta da metade da implementação começaram a surgir problemas de Overfitting com score local de 0.85-0.9 mas score no kaggle de apenas 0.75~. A partir daí comecei a pesquisar soluções em outros trabalhos e a modificar as variáveis. Por fim o modelo ficou com overfitting moderado e foi possível aumentar um pouco o score no kaggle.

Paralelo do problema ou Reutilização: O mais próximo para reutilização do modelo seria por exemplo uma brigada de bombeiros ou agentes de resgate utilizarem para verificar as chances de um certo grupo de passageiros de naufrágio desaparecidos terem sobrevivido ou não.