

Projeto Final - Turing Academy 2024

Análise, Limpeza e Treinamento de Modelos

O Projeto Final do TA tem como objetivo avaliar os conhecimentos adquiridos pelos trainees ao longo da duração do Turing Academy e garantir que todos possuam os conhecimentos necessários para adentrar nas áreas de foco do Turing.

O projeto deverá ser desenvolvido em grupos de 3 pessoas durante o período das férias de julho, tendo como **prazo de entrega máximo o dia 05/08** (05 de Agosto, segunda-feira).

Modelo de Entrega

O projeto consistirá no treinamento e na otimização de um modelo de Machine Learning à escolha do grupo, além da análise e tratamento da base de dados fornecida. A entrega consistirá nos seguintes:

- Um Jupyter notebook contendo a Análise Exploratória e Limpeza do dataset estudado
- Um (ou vários) notebooks contendo os (ou alguns dos) treinamentos realizados
- Um breve pdf (breve mesmo) contendo as decisões metodológicas adotadas, isto é: processos de limpeza, pré-processamento, avaliação, escolha de modelos e escolha de hiperparâmetros

Esses arquivos devem ser armazenados em um diretório no GitHub, cujo link de acesso deve ser enviado no seguinte formulário: [Envios \(GitHub\) - Projeto](#)

[Final](#). Nós recomendamos fortemente o uso do Git pelos integrantes do grupo para o manuseamento de arquivos do diretório.

Grupos

Os grupos serão formados por 3 integrantes e podem ser consultados na planilha a seguir: [Grupos - Projeto Final](#). Caso haja a necessidade de rever os grupos, por exemplo, existe um integrante do grupo que não vai poder participar do desenvolvimento do projeto, mandem mensagem para alguém da organização do TA.

Lembrem que os grupos não são de enfeite, trabalhem em grupo, é essencial para que possam aprender uns com os outros.

Etapas do Projeto

1. Análise de Dados e Limpeza

- Não é necessária a realização de uma Análise de Dados extensiva, mas é importante entender a correlação das features com a variável target

2. Treinamento Inicial e Avaliação

- Nessa etapa, podem ser treinados diversos modelos, sem alterações na etapa de processamento dos dados ou em hiperparâmetros, assim como testar múltiplas métricas

3. Treinamento Otimizado

- Nessa etapa, o objetivo é determinar o melhor modelo possível para a métrica determinada. Escolha um tipo de modelo e utilize diferentes técnicas para tentar melhorar o resultado (cuidado com overfitting). Você pode utilizar:

- Otimização de Hiperparâmetros
- Feature Engineering/Feature Selection
- Técnicas de Resampling
- Testar Estruturas de Redes Neurais
- etc...

4. Escrita do Relatório

- Aqui vocês devem colocar todas as decisões metodológicas utilizadas
- Além disso, incluam uma explicação do funcionamento básico do modelo escolhido e dos hiperparâmetros utilizados

5. (Bônus) Visualização

- Criem uma visualização a respeito dos dados trabalhados. Juntem diferentes features, personalizem estilos, fontes, etc, façam uma visualização diferenciada. Uma visualização bem feita às vezes vale mais do que 1000 comuns.

6. (Bônus) Deploy

- Após treinado um modelo final, realize o deploy do mesmo e crie uma aplicação que utilize o modelo treinado para realizar previsões.

Kaggle

A avaliação dos modelos treinados será realizada por meio da plataforma Kaggle. Além disso, o dataset do Projeto poderá ser encontrado na plataforma. Reparem que ela apresenta um leaderboard, no qual os grupos podem comparar seus resultados com os demais. Link do Kaggle: [Projeto Final - TA 2024 | Kaggle](#)