
Sessão 07

WiDS Recife Live Coding, 01/02/2020

Como serão os live codings?

- Sessões ao-vivo todos os sábados das 14h às 15h
 - Código e slides serão disponibilizados no nosso site
 - O objetivo é treinar para participar do [Datathon](#) em 2020
-

Roteiro

- Como usar o Pipeline() do Scikit-learn para “juntar” as etapas de treinamento e predição
- Enviando nossa primeira submissão para a competição

O problema

- O objetivo é criar um modelo que preveja a probabilidade do paciente sobreviver a partir dos dados das primeiras 24 horas dele na UTI do hospital
- A métrica utilizada para avaliação é a ROC AUC

Etapas para resolver o problema

1. Importar os dados
2. Explorar os dados
3. Treinar o modelo com o conjunto de treinamento
4. Avaliar o modelo com o conjunto de testes

Treinando um modelo

Treinando um modelo

- Vamos usar o `RandomForestClassifier()` com os parâmetros default, para criar um baseline
 - O conjunto de avaliação não possui labels, então vamos separar 40% do conjunto de treinamento para avaliar o modelo
-

Processamento dos dados

Processamento dos dados

- Precisamos fazer várias coisas:
 - Selecionar as features que vamos utilizar para a predição
 - Preencher os valores faltantes
 - Converter as features categóricas para one-hot-encoding
 - E por aí vai
 - Qual o problema disso?
 - Precisamos fazer as mesmas etapas para treinar o modelo e para fazer uma predição
-

Processamento dos dados

- Como o scikit pode nos ajudar?
 - Pipeline:
<https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>
 - Funciona como se nós juntássemos todas as etapas em um único objeto
 - Facilita muito na organização do código e na realização de experimentos
-

Pipelines

- Cada step precisa implementar a interface transform, e o último step pode ser um predictor, um transform ou os dois
 - Como criar uma pipeline
 - Primeiro definir os passos (steps)
 - Depois juntar todos os steps dentro da pipeline
-

Obrigada!

E até semana que vem!

Referências

- [1] <https://www.kaggle.com/c/widsdatathon2020/overview/description>