

Global Solutions – 1º Semestre 2022

Parabéns! Você acaba de ser contratado* para trabalhar no time de *Advanced Data Science* da Datafolha! Sua missão é empregar seus conhecimentos de Redes Neurais e Machine Learning para melhor entender as perspectivas eleitorais, por meio de dados.



Com a aproximação das eleições, o Datafolha se questionou se é possível construir um modelo preditivo para prever se um candidato ganhará as eleições antes mesmo que ela aconteça.

Este cenário é comum em qualquer país, no entanto, há uma ampla discussão interna na empresa sobre quais dados devem ser coletados, se a ideia funciona ou não, entre outras.

Para validar a ideia, você encontrou um dataset estrangeiro para construir uma prova de conceito. Se a ideia se tornar viável, você e seu time adaptarão o problema para o contexto brasileiro, ou seja, pensarão em novas features e coletarão os dados pertinentes.

Sua primeira missão no time de *Advanced Data Science* é explorar este conjunto de dados de eleições externas e construir um modelo preditivo que valide a ideia.

Veja mais informações a seguir.

Bom trabalho!

Dúvidas? Contate o Prof Michel

Base de dados

A base a ser utilizada será a mesma para as duas matérias e está junto com este documento, no Teams.

Disciplina: Machine Learning

PARTE 1. Entendimento do problema, dos dados e análises exploratórias

- Leia o dataset e reflita, para cada variável, qual é seu tipo e se ela seria relevante (ou não) para a detecção de quem ganharia a eleição. Pense também em outras variáveis que poderiam ser criadas a partir delas;
- Faça uma análise exploratória sobre os dados, narrando seus achados (ex: tamanho da base, tipos dos dados, porcentagem de dados faltantes, dados categóricos/numéricos, etc). Trace alguns gráficos que ajudem no entendimento dos dados fornecidos. Por exemplo: gráficos de correlação, de barras, distribuição de eleitos por idade ou gênero, etc. Façam pelo menos 3 gráficos diferentes seguindo boas práticas (no mínimo: título, nomes dos eixos, legenda).

PARTE 2. Modelagem preditiva: este candidato ganharia a eleição?

- Desenvolver modelos preditivos que sejam capazes de prever se um candidato ganha as eleições (*winner*) ou não. Foquem em implementar todo o processo de ciência de dados, desde a limpeza dos dados até a apresentação do resultado. Seu grupo deve entregar 2 modelos distintos, treinados em diferentes algoritmos de *machine learning*. **ATENÇÃO: para esta disciplina não valem modelos baseados em Redes Neurais! Escolha dentre aqueles que vimos em Machine Learning!**

=> Entregar um jupyter notebook contendo:

- Nome e RM dos alunos
- Códigos e análises da PARTE 1
- Códigos e análises da PARTE 2
- Os comentários devem ser feitos usando células markdown;

Disciplina: Redes Neurais

PARTE 1. Discussão do potencial da Inteligência Artificial na segurança de urnas eletrônicas

- As soluções de biometria ajudam a aumentar a segurança das aplicações. Um exemplo disso é a coleta de dados biométricos pelo Tribunal Superior Eleitoral, para melhor identificar os eleitores. Dados biométricos incluem impressões digitais, íris e rostos. Este é um campo de pesquisa amplo e com diversos resultados práticos. Faça uma busca para compreender como Redes Neurais e Deep Learning têm auxiliado a aprimorar o reconhecimento de dados biométricos.

PARTE 2. Modelagem preditiva: este candidato ganharia a eleição?

- Em *Machine Learning* você já explorou, limpou e modelou o problema utilizando 2 algoritmos clássicos de aprendizado de máquina. Aqui, você deve construir mais 3 modelos diferentes --- usando o mesmo conjunto de features --- todos baseados em Redes Neurais. As diferenças dos modelos estão na sua complexidade (Perceptron ou Redes Múltiplas Camadas?) e na arquitetura (ex: número de camadas, número de neurônios, funções de ativação, normalização, número de épocas, uso ou não de early stopping, escolha do otimizador, etc).
- Faça esta parte da implementação no mesmo jupyter notebook da matéria de ML, pois facilita sua vida e análise!

=> Entregar um relatório em PDF (máximo de 3 páginas, sem capa), contendo;

- Nome e RM dos alunos
- **PARTE 1:** resultados da sua pesquisa
- **PARTE 2:** para cada hiperparâmetro configurado na Rede Neural, justificar seu motivo de escolha; comente os resultados dos 5 modelos do ponto de vista de métrica de desempenho