

TÍTULO: *Ensaio sobre uso de ML para previsões de resultado de eleições municipais para vereador*

DEFINIÇÃO DO PROBLEMA



É viável o uso de Machine Learning para previsões eleitorais de candidatos a vereador?

Quais colunas preditoras que melhor explicam a possível eleição de um candidato?

RESULTADOS E PREVISÕES



Identificar, dentre candidaturas registradas para o cargo de vereador, quais candidatos têm maiores chances de serem eleitos, ou seja: fazer previsões dos eleitos.

Colunas preditoras: características pessoais do candidato, características do partido do candidato, desempenho do partido em pleitos anteriores, valores gastos na campanha, etc.

AQUISIÇÃO DE DADOS



Do repositório de dados do TSE usamos informações de sete bases em arquivo csv: candidaturas (dataset principal), coligações partidárias, bens de candidatos, quantitativo de vagas, receitas e despesas de campanha e resultados de eleições anteriores.

Informações de nomes dos municípios, distritos e subdistritos do IBGE (arquivo xls).

MODELAGEM



Temos um problema de classificação binária (eleito ou não eleito).

Logo, precisamos de algoritmos de Machine Learning supervisionados.

Iniciamos com o modelo Decision Tree e posteriormente testamos Random Forest, XGBoost, Catboost, Logistic Regression e um Ensemble por votação de 4 desses modelos.

AValiação DO MODELO



Temos classes muito desbalanceadas logo, a Acurácia não é uma métrica muito indicada.

Maior interesse está na previsão correta de positivos, portanto a métrica Recall é importante, sem descuidar da Precision. F1-score, que leva em consideração ambas, é uma boa métrica para avaliação em conjunto com dados da matriz de confusão

PREPARAÇÃO DOS DADOS



Tratamos dados ausentes tanto no dataset principal, como nas colunas agregadas posteriormente.

Imputação variada, de acordo com a necessidade de cada coluna.

Ajustamos, quando possível, colunas com informações inconsistentes.