

PROJETO DE APLICAÇÃO PARA O BOOTCAMP ENAP/2021

Tempestividade do julgamento de defesa/recurso das infrações ambientais pelo Ibama

Luiz Geraldo Santos Wolmer – TCU/SecexAgroAmbiental

Contexto

- Avaliação do desempenho do **processo sancionador ambiental** do Ibama pelo TCU.
 - Institucionalização da etapa da conciliação ambiental;
 - **Tempestividade da etapa do contencioso;**
 - Funcionamento da sistemática de notificação.
- Oportunidade de **conhecer o desempenho do processo, fatores que influenciam na tempestividade da etapa do contencioso**, e propor melhorias.

Solução

- Por meio de um modelo de ML de classificação, predizer se uma infração terá contencioso tempestivo, bem como conhecer fatores que influenciam a tempestividade dessa etapa no processo sancionador do Ibama.
- > Contencioso: análise e julgamento da defesa principal (1ª instância) e/ou do recurso (2ª instância)

Dados disponíveis

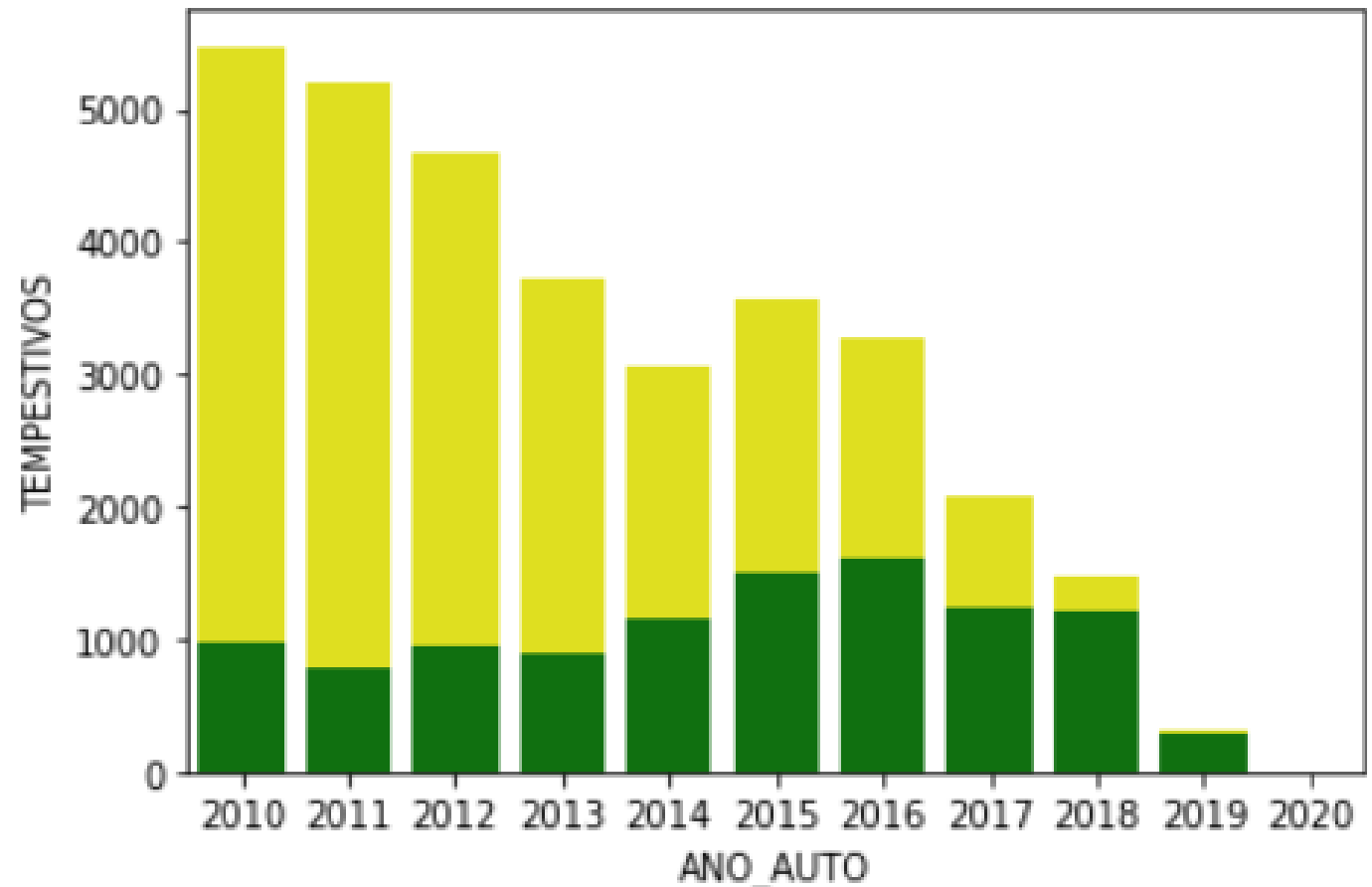
- **Inicialmente:** Dados abertos do Ibama (Autos de Infração*), desde 1981, contendo:
 - Infrator (nome, CPF/CNPJ); UF; Município;
 - Auto de infração (identificação, tipo de auto, tipo de infração, data e valor);
 - Datas dos julgamentos realizados;
 - Valor e data do pagamento;
 - Enquadramento legal, decisão administrativa e status do débito.
- * Disponível em <https://dados.gov.br/dataset/volume-de-julgamento-de-auto-de-infracao> para download por UF.
- **Usados no projeto:** dataset limpo e preparado, sem conter a identificação dos infratores (CPF/CNPJ e nome) e do auto de infração (NUMERO_AI): 285 mil registros

Etapas

- Limpeza e preparação de dados adicional -> autos de infração lavrados a partir de 2010, **em que houve julgamento de defesa e/ou recurso** (32.826 registros).
- Criação da variável dependente (**tempestividade**): diferença entre a data de início da instrução e a data do último julgamento realizado ≤ 360 dias.
- Treinamento de modelos de ML para prever a tempestividade, incluindo codificação (label/onehot), experimentação e escolha das variáveis independentes, bem como avaliação do grau de influência.

Registros por ano de autuação

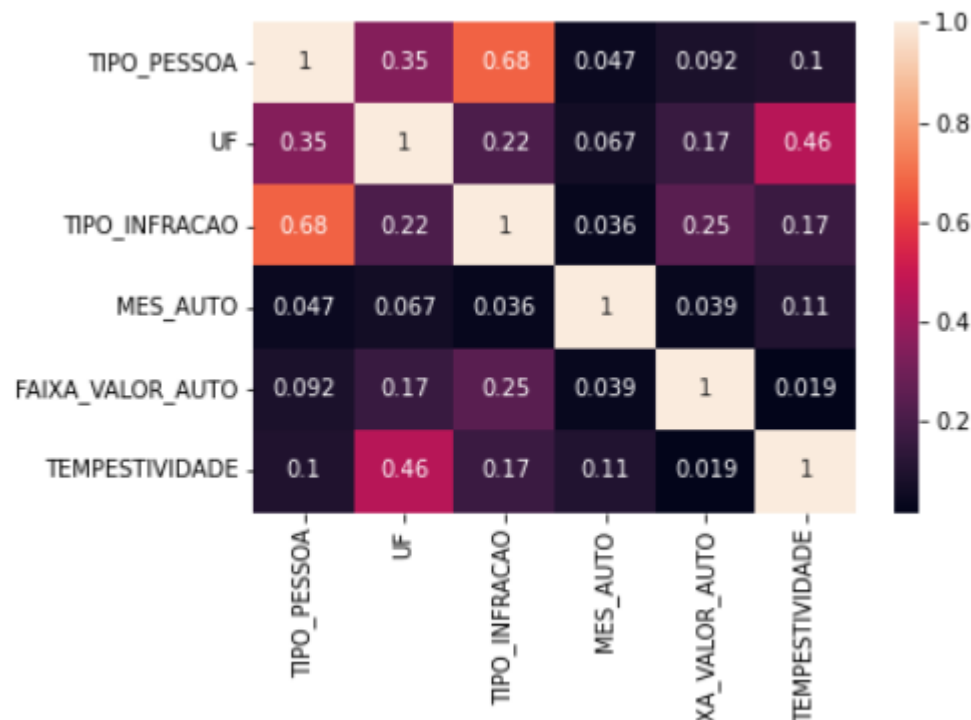
2010	5471
2011	5193
2012	4669
2013	3710
2014	3072
2015	3571
2016	3277
2017	2075
2018	1468
2019	318
2020	2



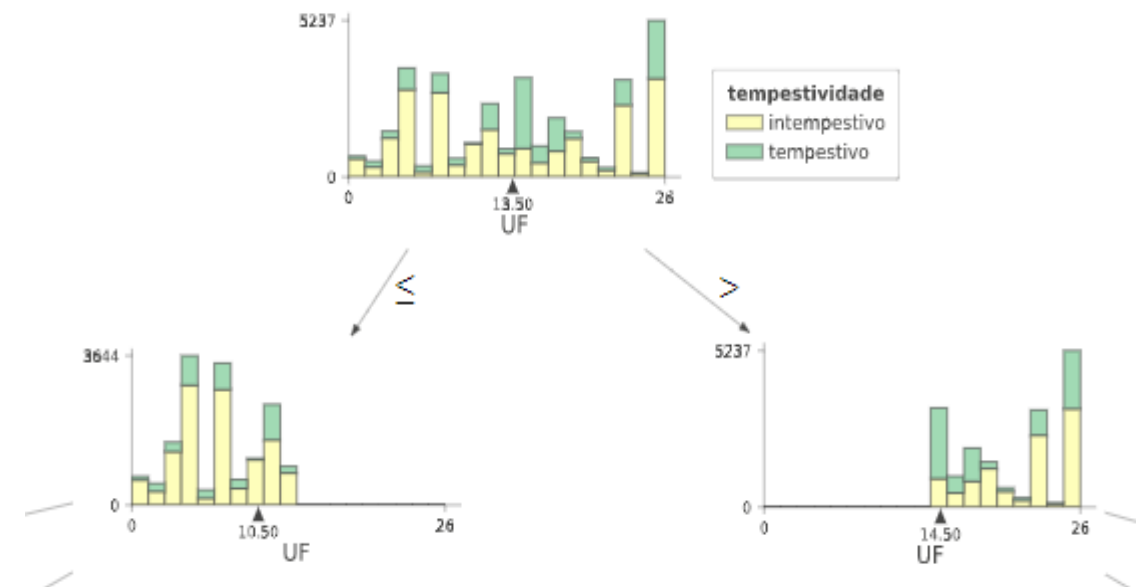
Resultados

- **Variáveis independentes finais:** TIPO_PESSOA, UF, TIPO_INFRACAO, MES_AUTO, FAIXA_VALOR_AUTO

V de Cramer



Topo da árvore (Dtreeviz)



Resultados

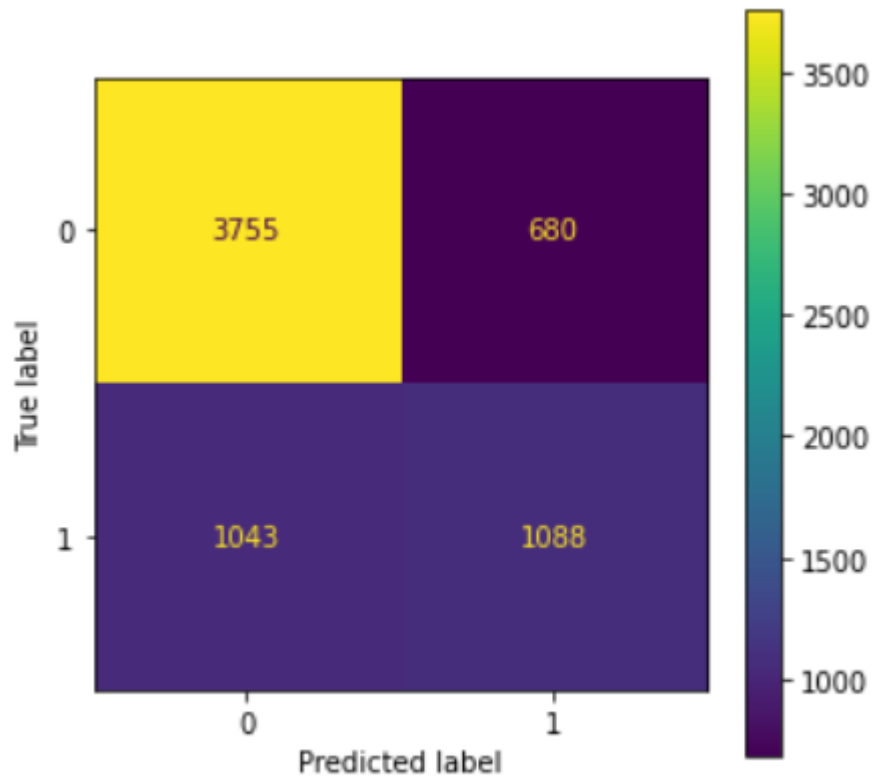
- Baseline = 68% intempestivos / 32% tempestivos
- Testes prévios: labelencoding, valor_auto normalizado, SMOTE, Catboost.
- Resultados para modelos isolados após onehotencoding:

Modelo	Acurácia	F1
Reg-Log	0,748	0,732
KNN	0,738	0,731
Árvore de decisão	0,744	0,731
Random Forest	0,733	0,669
XGBoost	0,751	0,728

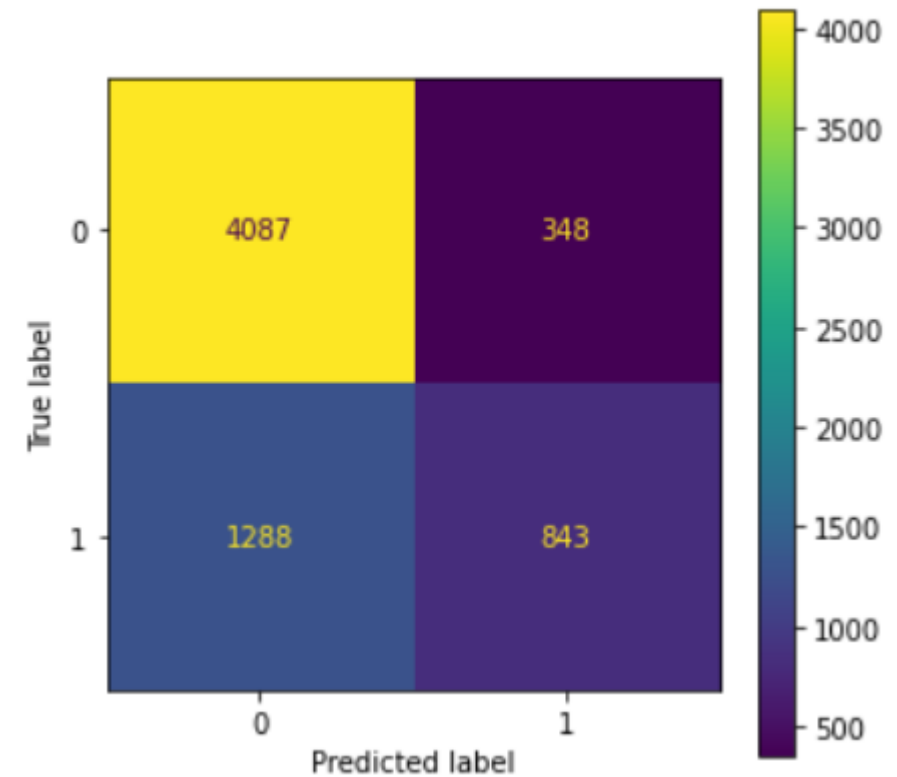
Resultados

Confusion Matrix

KNN



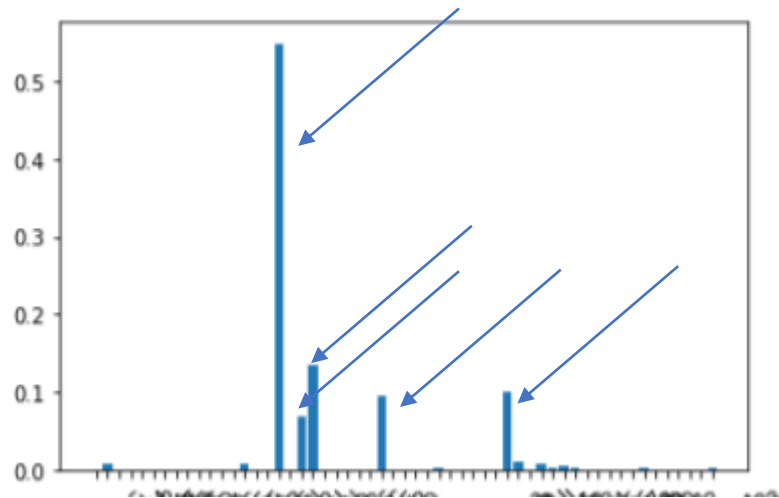
XGBoost



Resultados

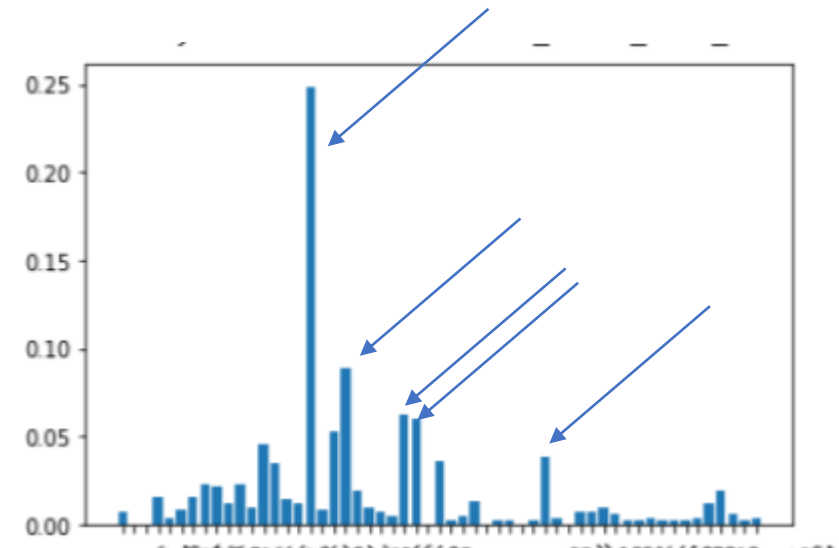
Feature Importance

Árvore de Decisão



0.54807 UF_PB
0.06833 UF_PI
0.13588 UF_PR
0.09626 UF_SC
0.10002 TIPO_INFRACAO_Outras

XGBoost



0.24858 UF_PB
0.08921 UF_PR
0.06207 UF_RS
0.06032 UF_SC
0.03851 TIPO_INFRACAO_Outras

Resultados

- Modelos testados com gridsearch/cross validation:

Modelo	Melhor resultado
KNN n_neighbors = range(5, 10, 15) weights = ['uniform', 'distance'] metric = ['None', 'euclidean', 'manhattan']	Best: 0.69033 using {'metric': 'euclidean', 'n_neighbors': 5, 'weights': 'distance'}
Random Forest 'max_depth': [5, 8, 'None'], 'max_features': ['log2', 'None'], 'min_samples_leaf': [1, 2], 'min_samples_split': [2, 5], 'n_estimators': [100, 300]	Best: 0.60101 using {'max_depth': 8, 'max_features': 'log2', 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 100}
XGBoost 'colsample_bytree': [1, 2], 'gamma': [0, 1], 'learning_rate': [0.01, 0.1], 'max_depth': range(3, 6), 'n_estimators': [50, 100]	Best: 0.68183 using {'colsample_bytree': 1, 'gamma': 1, 'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 100}

Próximos passos

- Minerar dados recebidos do Ibama para depurar a base atual e incluir mais variáveis das **infrações**.
- Incluir variáveis qualificadoras dos **infratores**, além da classificação PF/PJ, a partir de outras bases de dados.
- Ampliar predição para **outras circunstâncias** do processo sancionador:
Eficácia do processo de **notificação**.
Efeitos da etapa de **conciliação**.

PROJETO DE APLICAÇÃO PARA O BOOTCAMP ENAP/2021

Tempestividade do julgamento de defesa/recurso das infrações ambientais pelo Ibama

Luiz Geraldo Santos Wolmer – TCU/SecexAgroAmbiental