

# Predição de Reincidência Criminal

## Dataset

O National Corrections Reporting Program coleta informações administrativas sobre admissões e libertações de prisões, além de populações de custódia ao longo do ano. Dados demográficos, infrações por condenação, duração da sentença, tempo mínimo a ser atendido, tempo de prisão creditada, tipo de admissão e liberação, e tempo de espera são coletados dos registros individuais de prisioneiros. Os dados são publicados no portal de dados do governo federal.

## Modelos

Quatro modelos foram escolhidos para atacar o problema:

- Naive Bayes
- Random Forest
- LightGBM
- XGBoost

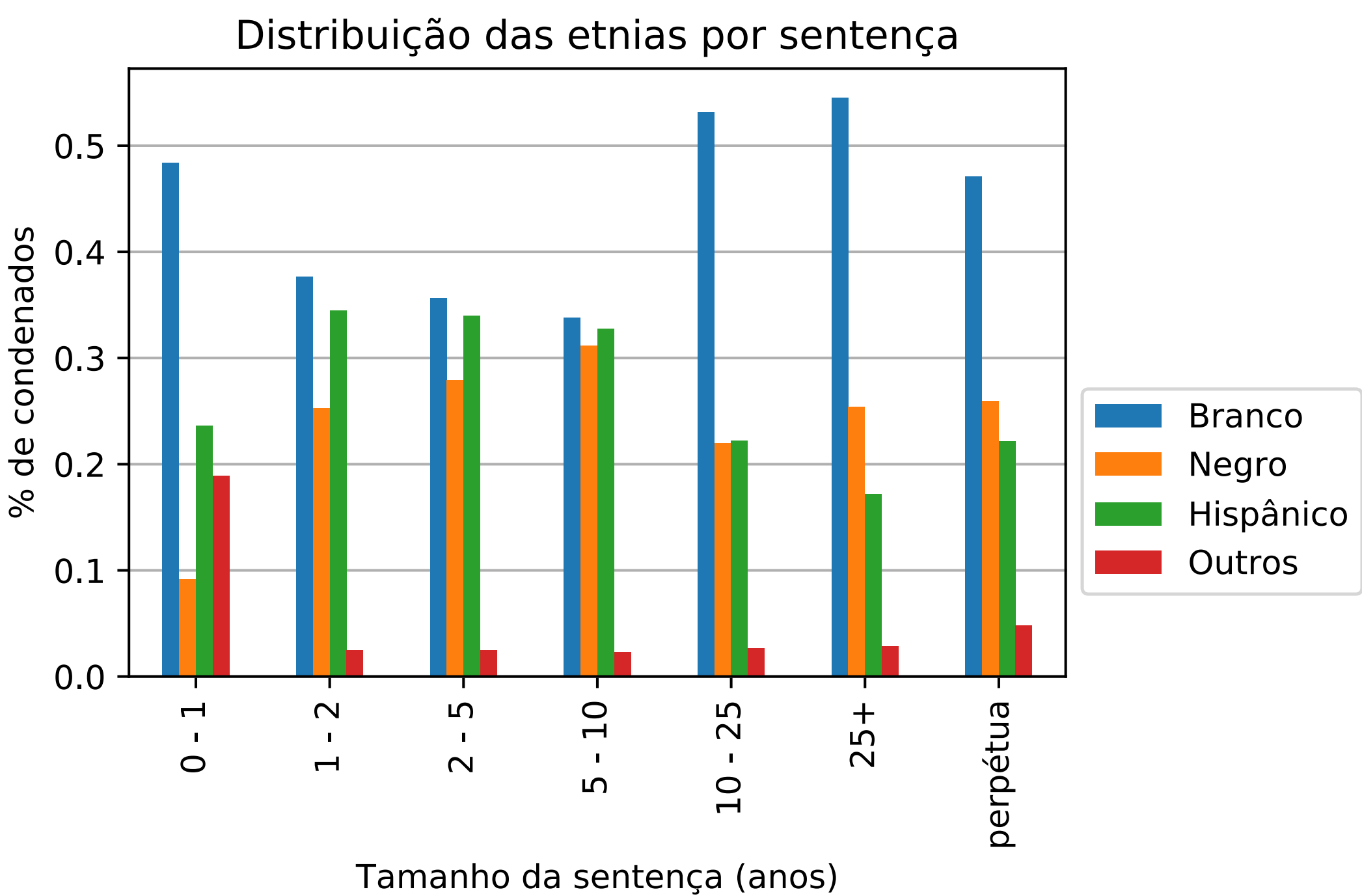
Outros modelos foram desconsiderados devido à limitações da capacidade computacional disponível.

## Resultados

F1 Score				
Tamanho do dataset	Naive Bayes	Random Forest	LightGBM	XGBoost
2 Milhões	66.0%	62.9%	61.6%	62.5%
4 Milhões	71.1%	69.9%	71.1%	-
6 Milhões	68.8%	66.0%	69.9%	-

## Influência da Etnia nas sentenças

Apesar da distribuição das etnias nas sentenças não indicar a princípio um viés de raça, os modelos indicam uma relevância considerável desta característica nos dados.



F1 Score dos modelos				
	Naive Bayes	Random Forest	LightGBM	XGBoost
Sem etnia	66.0%	62.9%	61.6%	62.5%
Com etnia	66.3%	64.4%	66.3%	63.8%
Melhoria	0.3%	1.5%	4.7%	1.3%