Determinantes das decisões judiciais em audiências de custódia

José de Jesus Filho e Julio Trecenti

VII EPED - UEFS, Feira de Santana, 24 de agosto de 2017

Contexto

- Estudo sobre audiências de custódia em São Paulo
- ► Abordagem utiliza jurimetria, estatística aplicada ao direito
- No TJSP, aproximadamente 88% das decisões são desfavoráveis ao autor
 - ▶ 89% em Habeas Corpus (HC) da defesa
 - ▶ 65% em Mandados de Segurança (MS) do Ministério Público



Questões norteadoras

- É possível predizer o resultado com alta taxa de acerto?
- Quais são as variáveis que mais influenciam no resultado?

Metodologia em 3 passos

Download automatizado de decisões do TJSP



Metodologia em 3 passos

▶ Parse. Aplicação de técnicas de mineração de texto para extrair variáveis e montar uma base de dados



Metodologia em 3 passos

► **Model**. Aplicação de modelos estatísticos para predizer os resultados.



Detalhes sobre o Download

- Pesquisa de jurisprudência do TJSP
- Palavra-chave: "audiência de custódia"
- ▶ Download de ~3900 decisões
- ► Filtro: apenas HCs e MSs
- Necessidade de quebrar CAPTCHAs para baixar arquivos PDF



 Processo automatizado a partir dos pacotes esaj e tjsp, construídos com base no software estatístico R

Detalhes sobre o Parse

- Aplicação de expressões regulares para extração dos resultados e informações do processo
- Exemplo: extração da decisão

```
"(denega.*|denaga.*|.*deferi.*|conced.*)"
```

4 Grupos de variáveis

Atores do processo

► Impetrante: MP ou defesa

► Câmara julgadora: 01 a 16.

Gravidade da pena

Gravidade: baixa, média, alta

Primariedade e antecedentes

Primariedade: sim ou não

Características do processo

- ▶ Garantia de aplicação da lei, art. 312 do CPP: sim ou não
- Conveniência da instrução criminal: sim ou não
- Manutenção da ordem pública: sim ou não

Detalhes sobre modelagem

- Modelo estatístico:
 - busca criar uma história sobre os dados
 - simplifica a realidade e assume que existem erros
 - busca minimizar esses erros
- **Exemplo**: árvore de decisão



▶ Modelo trivial: chutar sempre a categoria mais frequente

Detalhes sobre modelagem

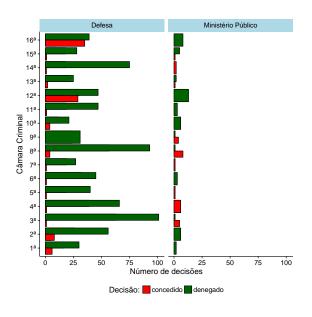
► Teste / treino



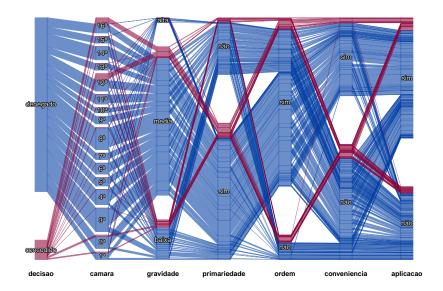
- Modelo é ajustado na base de treino
- Comparamos os resultados do modelo aplicado à base de teste com os resultados observados

Resultados

Descritiva

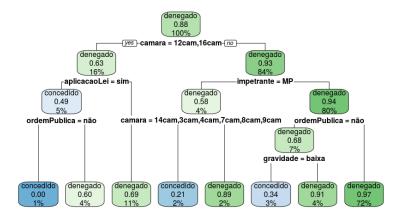


Descritiva



Árvore de decisão

Resultados: acurácia 90%, sensibilidade 29%, especificidade 98%.



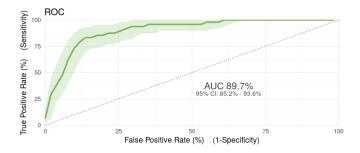
Problemas

- ► Apesar da taxa de acerto alta, sensibilidade é baixa
- Definição de sensibilidade, ou taxa de verdadeiros positivos:

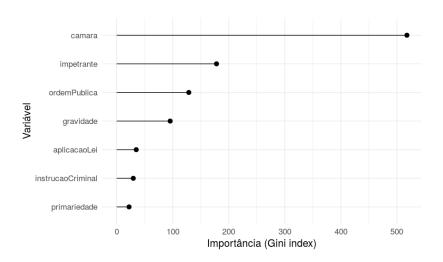
$$\textit{TPR} = \frac{(\text{chutei concedido e acertei})}{(\text{chutei concedido e acertei}) + (\text{chutei denegado e errei})}$$

Outros modelos

- Aplicamos modelos mais sofisticados, como XGBoost, Florestas aleatórias e Redes Bayesianas
- Melhores resultados obtidos com florestas aleatórias
 - ▶ 84% de acerto
 - ▶ 84% de especificidade
 - ▶ 83% de sensibilidade
- Curva ROC



Importância das variáveis



Conclusões

- É possível predizer decisões judiciais com base em informações do processo
 - ▶ No nosso caso, é importante trabalhar a sensibilidade
- As variáveis mais importantes são relacionadas ao grupo de atores: câmara e tipo de impetrante.
- A jurimetria é essencial para compreender o funcionamento do Direito.

Obrigado!

- José de Jesus Filho
 - ▶ jjesusfilho@gmail.com
 - http://www.josejesus.info
- Julio Trecenti
 - ▶ jtrecenti@abj.org.br
 - http://abj.org.br

