

Análise quantitativa de processos criminais

José de Jesus Filho

Janeiro de 2018

E-mail: jjesusfilho@gmail.com

Telefone: (11)98522-0210

CNPJ: 23.268.449/0001-30

Web: josejesus.info

Endereço: Rua Waldemar Adelino da Silva, 30

São Paulo - SP CEP 02929-020

Ementa

O curso irá introduzir a profissionais e estudantes de direito uma metodologia para obter e analisar dados dos processos judiciais criminais dos tribunais estaduais e federais. Inicialmente, serão introduzidos aos fundamentos teóricos do realismo jurídico. Os participantes irão familiarizar-se com a linguagem de programação R e, num segundo momento, irão aprender técnicas de raspagem de dados na web (web scraping), processamento de linguagem natural (natural language processing-NLP), limpeza e organização de base de dados. Na última etapa, os participantes aprenderão as técnicas de análise descritiva e preditiva de dados jurídicos.

O curso será eminentemente prático, as bases teóricas somente serão fornecidas na medida em que sirvam para elucidar conceitos imprescindíveis para a compreensão das operações realizadas.

Programa do curso

Aspectos teóricos

1. Paradigmas de interpretação do comportamento judicial
2. Diferenças e complementariedade na pesquisa quanti e quali no direito
3. Introdução à jurimetria
4. Introdução à criminometria

Introdução ao R

1. Objetos no R: números e vetores
2. Outros objetos no R: fatores no R

3. Outros objetos no R: caracteres
4. Outros objetos no R: matrizes
5. Outros objetos no R: listas e data frames
6. Importando dados para o R
7. Exportando dados do R
8. Execução condicionada: ifelse
9. Execução condicionada: loops: for, repeat, while
10. Noções de distribuições probabilísticas
11. Estatística básica: média, mediana, moda, variância, desvio padrão, erro padrão
12. Criando funções no R

Coletando dados dos tribunais de justiça

1. Noções de html, xml, json e css
2. Protocolo http
3. Noção de ajax
4. Requisitando dados da web: httr e rvest
5. Extraíndo conteúdo da página: rvest, xml2, css e xpath
6. Extraíndo conteúdos dinâmicos: RSelenium
7. Baixando arquivos da web: pdf, csv, excel

Manuseio de textos

1. Introdução ao processamento de linguagem natural
2. Conversão e importação dos textos
3. Expressões regulares
4. Extraíndo informações relevantes das peças processuais
5. Tokenização, lemanenização, kwic e outras técnicas de manuseio de textos
6. Estruturando textos em tabelas, planilhas e matrizes

Manuseio da base de dados

1. Introdução ao tidyverse
2. Manuseio de dados com dplyr
3. Programação funcional no R com purrr
4. Manuseio de datas: lubridate
5. Manuseio de fatores: forcats
6. Deixando a base pronta para análise

Introdução ao Aprendizado de Máquina (Machine Learning) no R

1. Noção de Machine Learning
2. Aprendizado supervisionado vs aprendizado não supervisionado

3. Overfitting vs underfitting
4. Regularização
5. Noção de entropia
6. Regressão logística
7. Árvores de decisão
8. Floresta aleatória
9. Boosting: Adaboost, Gradient boosting, Extreme Boosting
10. Outros algoritmos: SVM, naive bayes, knn

Rodando modelos de predição das decisões judiciais

1. Dividir para conquistar: criando as bases de treinamento e teste
2. Estabelecendo hiperparâmetros
3. Validação cruzada
4. Rodando o modelo com regressão logística
5. Rodando o modelo com Floresta aleatória
6. Rodando o modelo com boosting
7. Calibrando os modelos

Avaliando o desempenho dos modelos rodados

1. Comparando os modelos
2. Verificando as métricas de desempenho de cada modelo: matriz de confusão, acurácia, sensibilidade, especificidade,
3. Gráfico de Kolomogorov Smirnov
4. Área sob a curva: AUC-ROC
5. Coeficiente de GINI

Realizando predições

1. Aplicando a modelo sobre a base teste
2. Predizendo valores individuais
3. Interpretando as probabilidades

Material utilizado no curso

Além do software gratuito e de seu IDE RStudio, o palestrante oferecerá material de estudo e de exercícios.

Duração do curso

O curso durará no mínimo 40 horas, podendo chegar a 45 horas, a depender do andamento na cobertura do programa.

Custos

Definir