

Desafio Técnico JuMP — Cientista de Dados

Esta é uma proposta de desafio que será utilizada no processo de seleção do projeto **JuMP** para contratação de um bolsista para atuar como *cientista de dados*. O Projeto trabalha com o conceito de mineração de processos, portanto, o objetivo desse desafio é manipular os dados dos processos judiciais anonimizados e extrair insights. A seguir você terá mais detalhes do problema e dos requisitos para a realização do desafio. Você **não** será eliminado por deixar de cumprir todos os requisitos, mas tente fazer o máximo da melhor forma possível.

Conceitos importantes

É importante relacionar alguns conceitos de mineração de processos para a nossa realidade, uma vez que a palavra “processo judicial” da área se refere a um “Caso” específico. Onde cada caso possui diversas atividades em uma sequência no tempo, e essas atividades podem possuir recursos atrelados a elas. Veja abaixo a tabela de relações:

Mineração de processos	Judiciário ou Teste técnico
Event Log	Processos judiciais ou Dataset
Case	Processo Judicial
Activity	Movimento
Timestamp	Data Final (movimento não tem início e fim, então tanto faz)
Resource	Documentos/Complementos

Os processos judiciais possuem uma natureza altamente não estruturada, sendo compostos por uma variedade de movimentos que ocorrem em sequências diversas e imprevisíveis. Essa complexidade processual é agravada pelo fato de que movimentos insignificantes, muitas vezes repetitivos, surgem com alta frequência, enquanto movimentos realmente significativos podem ser raros ou até ausentes, o que dificulta sua identificação e análise. Por exemplo, movimentos como "Publicação", "Decurso de Prazo", "Conclusão" e "Disponibilização no Diário da Justiça Eletrônico" são frequentemente repetidos, e não são úteis, pois não contribuem de maneira significativa para a análise.

Uma abordagem eficaz para destacar movimentos importantes, mas de baixa frequência, é agrupá-los com movimentos similares ou relacionados. Esse agrupamento permite uma análise com generalização desses movimentos e facilita a identificação de padrões relevantes. Para realizar essa tarefa, utilizamos a Tabela de Padronização de Unidades (TPU) do CNJ, que pode ser acessada a partir do [arquivo especificado](#).

Abaixo, apresentamos uma lista de movimentos que devem ser agrupados com base em seus respectivos identificadores. Esse processo de agrupamento reflete uma generalização dos movimentos em categorias mais amplas.

Identificador	Nome do movimento
3	Decisão
11009	Despacho
193	Julgamento
14092	Voto
12430, 228	Arquivamento

Por outro lado, é fundamental fornecer um contexto mais claro sobre a execução de determinadas atividades, já que essas atividades podem ocorrer em diferentes momentos ao longo do processo judicial. Por exemplo, o movimento de "petição" pode ter vários contextos, como petição inicial, petição de contestação, ou petição de juntada de documentos, entre outros, cada um representando uma fase distinta do processo. Para capturar essa diversidade de contextos, o dataset inclui duas colunas específicas — "documento" e "complemento" — que oferecem informações adicionais, permitindo uma especialização mais precisa de alguns desses movimentos. Para realizar essa tarefa, utilizamos a Tabela de Padronização de Unidades (TPU) do CNJ, que pode ser acessada a partir do [arquivo especificado](#). Abaixo, apresentamos uma lista de movimentos que devem ser especializados(especialize apenas os que estão na tabela, atenção para os identificadores):

Identificador	Nome do movimento
85,12271	Petição

60	Expedição de Documento
11010	Mero Expediente
106,985	Mandado
970	Audiência

Desafios

Como cientista de dados no projeto JuMP, é crucial que você tenha a capacidade de simplificar a visualização dos processos judiciais por meio de técnicas avançadas de descoberta de modelo de processo (mineração de processos), além de extrair insights significativos a partir dos dados. Você foi encarregado de realizar uma análise detalhada dos movimentos presentes no dataset de duas unidades judiciárias e compará-las. O objetivo é aprimorar o entendimento contextual de determinados movimentos (generalização/especialização) e elaborar visualizações comparativas sobre a performance e frequência da execução dos processos judiciais. Abaixo estão os desafios que essa tarefa envolve:

1. (Opcional) Remover os movimentos insignificantes
2. (Opcional) Agrupar os movimentos de acordo com a árvore CNJ
3. Especializar os movimentos por documentos e/ou complementos
4. (Opcional) Trazer uma análise descritiva do melhor fluxo gerado
5. Descrever as estatísticas dos dados sob a perspectiva dos movimentos, complementos e documentos e suas relações. Seja criativo e forneça visualizações que comuniquem essas informações para os usuários e pesquisadores do JuMP. Foco na comparação do perfil estatístico das duas unidades.

Ferramental

Para realizar essa tarefa, você deve realizar o download do [dataset anonimizado](#), e utilizar a biblioteca [PM4PY](#), em qualquer versão, para criar o modelo processual utilizando o algoritmo Heuristics Miner e exibir sua visualização.

Requisitos Desejáveis

- Utilizar o [Polars](#) em vez de Pandas para a manipulação dos dados; (OPCIONAL)
- Realizar o máximo de requisitos necessários que você conseguir;
- Trazer várias visualizações ou insights a partir dos dados obtidos;

Considerações

- Fique à vontade para entregar os scripts em qualquer plataforma (jupyter notebook, Colab, Briefer, tanto faz).
- **Implemente o que for possível, priorize o que na sua opinião irá trazer mais valor e melhore o seu projeto o máximo que conseguir;**
- Vamos analisar tudo que você fez e até onde você conseguiu chegar, então documente e não deixe de entregar o projeto, mesmo que esteja incompleto;

Boa sorte!

CONSIDERAÇÕES FINAIS:

Entrega parcial: 22/08/2024 23:59 (opcional)

Data de entrega: 25/08/2024 23:59

Forma de entrega: Enviar e-mail para jdsc@cin.ufpe.br e tsa2@cin.ufpe.br contendo o link ou arquivo que possui os resultados do teste técnico.