



## **Plano de Ensino** **Clínica 2**

### **Sumário**

O cliente é a e-xyon, uma lawtech com centenas de colaboradores fundada em 2001 que oferece ferramentas de (RPA) robôs de automação, workflow e IA (Inteligência Artificial) integrados aos seus softwares. Os alunos serão alocados em 8 grupos que competirão para desenvolver a versão mais satisfatória de um sistema em Python que recebe como input metadados e inteiro teor de petições iniciais protocoladas na justiça estadual e identifica os processos cuja causa raiz é crédito consignado, bem como o valor de indenização solicitado e produz dados estatísticos sobre o conjunto das petições.

### **Desafio**

Elaborar um sistema que seja capaz de ler um conjunto qualquer de metadados de processos da justiça estadual e inteiro teor das respectivas petições iniciais para a) identificar os processos cuja razão central para que o autor tenha ajuizado a ação é uma disputa sobre crédito consignado e b) os valores exigidos a título de indenização pelo autor do processo. A partir da extração dessas informações, o programa deve produzir dados estatísticos, inclusive com representações gráficas, de fatores associados aos valores de indenização pedidos, a diferentes empresas processadas e outros fatores.

Um relatório será entregue pelos grupos, porém não será o produto principal e não será lido pela banca avaliadora.

Os alunos receberão [um arquivo](#) em formato csv com 19800 linhas contendo metadados de cada processo, incluindo partes do texto da petição inicial.

Os alunos devem usar esse conjunto de documentos para planejar, desenvolver e testar seu sistema. O sistema de cada grupo será testado usando um conjunto separado de documentos antes da banca avaliadora.

Ao receber um conjunto de arquivo csv e arquivos txt, o sistema deverá produzir:

1) estatística descritiva de pedidos na petição inicial sobre indenização por danos materiais e morais quando se trata da causa raiz crédito consignado.

Isso pode incluir, por exemplo, o valor médio em reais que aparece nas petições iniciais quando se trata de pedido sobre danos materiais;

2) modelo inferencial indicando características do caso e respectiva petição inicial associadas aos processos sobre crédito consignado;

3) análises e soluções escolhidas pela equipe considerando as demandas e necessidades da advocacia na prática do direito civil.

Cada grupo deverá considerar especialmente o item 3 como forma de empregar criatividade na proposta daquilo que o sistema irá entregar. Entrevistas com profissionais da advocacia para coleta de informações são encorajadas.

Exemplos de soluções para atender ao desafio 3:

a) um chatbot que seja capaz de responder a um advogado perguntas sobre o perfil de processos ajuizados sobre diferentes causas raiz na justiça estadual contra instituições financeiras, incluindo, mas não limitado a, crédito consignado;

b) um dashboard interativo mostrando o comportamento de variáveis importantes sobre os processos ou o os elementos textuais das petições iniciais ao longo do tempo ou sobre partes processadas diferentes

No dia 6 de novembro, os grupos devem submeter pelo Blackboard o código fonte de seu programa. O arquivo comprimido em formato ZIP deve indicar o número do grupo em seu nome e conter o código fonte - em arquivo .ipynb - e arquivos acessórios desejados pelo grupo.

O código fonte deve conter o comentário "LINHA DE SELEÇÃO DO INPUT" logo acima da linha que realiza a importação do arquivo .csv usado na análise. Essa linha será alterada pelo professor para importar um .csv de teste contendo banco de dados em formato similar àquele usado pelos alunos para desenvolver e testar seu programa. O arquivo terá as mesmas características (nomes das colunas, tipo dos dados) daquele disponibilizado aos grupos no início da Clínica 2.

Os grupos podem utilizar qualquer biblioteca em seu código. Dependendo da biblioteca, assumem a responsabilidade pelo caso de a sua instalação não ser possível, após várias tentativas, pelo professor antes de rodar o programa.

Ao ser executado, o software do grupo deve salvar:

1) um arquivo output.xlsx no mesmo local do arquivo do código fonte. Esse arquivo é uma planilha com a lista dos processos selecionados, após passar pelos filtros (ou seja, obter apenas os processos nos quais a petição inicial indica a causa raiz crédito consignado) e enriquecimento com informações que o grupo considerou analisar, incluindo novas informações coletadas a partir dos próprios dados ou fontes externas. A base de dados deve conter pelo menos uma coluna "id", contendo os números únicos de identificação dos documentos que fazem parte do escopo, mas pode incluir outras colunas que o grupo desejar.

2) um arquivo relatorio.xlsx no mesmo local do arquivo do código fonte. Esse arquivo deve conter:

- a) A porcentagem de processos, dentre todos os oferecidos como input, nos quais a petição inicial indica a causa raiz crédito consignado.
- b) A média do valor em reais dos pedidos de condenação por danos materiais feitos nas petições iniciais sobre crédito consignado.
- c) A média do valor em reais dos pedidos de condenação por danos morais feitos nas petições iniciais sobre crédito consignado.

Caso o código fonte realize requests para APIs que exijam chave, o grupo deve deixar a chave no código submetido e deverá cancelá-la após o dia 11/11.

## **Calendário**

12/8, 12-14h – Apresentação do Desafio, informações sobre o formato, atuação dos tutores, descrição dos dados

19/8, 12-14h – Aula com Bruno Oliveira sobre uso de LLMs para extração de features

26/8, 12-14h - Aula com Graziela Tonin sobre método ágil

2/9, 12-14h – Aula com Guilherme Almeida sobre filtragem de dados textuais

9/9, 12-14h - Aula com Helena Funari sobre expressão regular

11/9, 12h30-14h - Horário de atendimento com Helena Funari

16/9, 12-14h - Aula com Henrique Wang sobre clusterização

23/9, 12-14h - Aula com Mauro Sampaio da e-xyon sobre a atuação da empresa no mercado jurídico e os desafios trazidos pela IA

7/10, 12-14h - Aula com Roberto Maruyama da e-xyon sobre a identificação de causas raiz em processos

9/10, 12h30-14h - Horário de atendimento com Henrique Wang

21/10, 12-14h - Aula com Julio Trecenti sobre modelagem preditiva

23/10, 12h30-14h - Horário de atendimento com Bruno Oliveira

### **28/10, 12-14h - Prova Individual**

4/11, 12-14h - Aula com Ivar Hartmann com recomendações sobre a apresentação nas bancas

6/11, 12h30-14h - Horário de atendimento com Julio Trecenti

**6/11, até 23:59 - submissão do relatório e código fonte do sistema no Blackboard**

**11/11, Nos slots 10h15-11h30; 12h30-13h45; 15h45-17h45 – Bancas finais - Apresentação da solução por cada grupo**

### **Avaliação**

A nota de cada aluno será composta por

\*Prova: 35% (individual): Prova individual e sem consulta com 12 questões múltipla escolha em 28/10, 12-14h. As questões serão elaboradas com base nas aulas com os tutores

\*Relatório e código-fonte: 20% (grupo): Documentos submetidos pelo grupo em 9/6. A nota levará em consideração o cumprimento dos requisitos básicos listados acima para o código-fonte. O relatório deverá ter entre 2 mil e 3 mil palavras e conter 1) Descrição das etapas de tratamento dos dados; 2) Descrição da etapa de análise estatística descritiva, mostrando gráficos produzidos pelo sistema; 3) Descrição da metodologia utilizada para elaborar os modelos inferenciais e seus resultados.

\*Bancas: 45% (grupo): Performance do grupo nas bancas. 2 finalistas: nota 10. 2 semifinalistas: nota 8. 4 grupos não classificados para semifinal: nota 5. A nota será atribuída ao grupo. Qualquer componente do grupo ausente em uma das bancas do grupo terá sua nota referente às bancas zerada.

## **Bancas**

No dia 11/11 serão realizadas 3 rodadas de bancas. Cada grupo participará de uma banca com apenas um outro grupo. A banca decidirá qual dos dois grupos apresentou uma solução melhor.

O tempo da banca será utilizado para 10 minutos de apresentação e tempo de arguição conforme a rodada - 10 minutos na primeira, 15 na segunda, 20 na terceira. A apresentação será feita exclusivamente pelo componente do grupo sorteado pelo professor no momento da apresentação.

O chaveamento será decidido com base nos relatórios entregues pelos grupos no dia 6/11. O professor fará um esforço para alocar os grupos nas bancas de modo que os grupos que apresentaram os melhores relatórios não realizem banca juntos na primeira rodada.

10h15-11h30: 4 bancas com 2 grupos em cada uma

Os grupos escolhidos por sua melhor solução em cada uma das 4 bancas realizarão uma segunda rodada de bancas

12h30-13h45: 2 bancas com 2 grupos em cada uma

Os grupos escolhidos por sua melhor solução em cada uma das 2 bancas realizarão uma banca final. Não haverá banca com os grupos com a 3a e 4a melhores colocações. Isso será definido com base na mediana da nota dos integrantes de cada grupo na prova individual.

15h45-17h45: Uma banca final

## **Premiação**

1a colocação: R\$ 10 mil

2a colocação: R\$ 5 mil

3a colocação: Menção honrosa

**Para decidir qual grupo apresentou a melhor solução para o desafio, a banca aplicará, de forma subjetiva, os seguintes critérios:**

-Qualidade: avaliada a partir do relatório entregue à banca pelo professor da Clínica após testes do programa de cada grupo submetido em 6/11. A qualidade diz respeito à precisão da estatística \*descritiva\* (item 1 de 3) que a solução de cada grupo deve produzir sobre os documentos válidos

-Valor: avaliado a partir do conjunto global de funcionalidades do programa apresentado pelos grupos na banca. O valor do programa para escritórios de advocacia ou outras instituições públicas ou privadas com atuação no direito brasileiro, inclusive departamentos jurídicos de empresas, será determinado pela banca a partir da identificação de demandas ou necessidades comuns nessa atuação que tais instituições possuem e o programa apresentado atende de maneira satisfatória.

-Inovação: avaliado a partir do conjunto global de funcionalidades do programa apresentado pelos grupos na banca. O nível de inovação oferecido pelo programa apresentado quando comparado com o outro programa na mesma banca ou quando comparado com soluções existentes no mercado jurídico que buscam atender demandas similares.

-Eficiência: avaliado a partir do conjunto global de funcionalidades e custos do programa apresentado pelos grupos na banca. Esse critério envolve a comparação do custo-benefício de cada programa apresentado à banca, considerando os custos de operação e, especialmente, custos atrelados ao uso contínuo do programa como requisições a APIs de grandes modelos de linguagem (ex.: ChatGPT).