BOLSA DE INICIAÇÃO CIENTÌFICA – PROCESSO SELETIVO

O Escritório de Apoio à Pesquisa comunica a abertura de processo seletivo para indicação de um bolsista na modalidade Iniciação Científica para atuar sob orientação do Prof. Dr. Ildeberto Aparecido Rodello no âmbito do projeto "Implementação de uma Rede Neural Multicamadas para a Classificação e Análise de Decisões Judiciais" (ANEXO I), vinculado ao edital de Apoio a Projetos que Façam Uso de Sistemas Digitais Inteligentes (Portaria PRP nº 668, de 17 de outubro de 2018).

1. REQUISITOS

Vaga exclusiva para alunos regularmente matriculados nos cursos de graduação da USP, cumprindo os seguintes requisitos:

a) estar regularmente matriculado em curso da área de tecnologia da Informação ou Computação, cursando o 3º ano/6º semestre ou posterior, com conhecimentos avançados em lógica de programação, linguagem de programação Python e suas bibliotecas NLTK e SciKit Learn. São desejáveis conhecimentos sobre inteligência artificial e redes neurais.

b) não possuir vínculo empregatício ou receber simultaneamente outra bolsa de qualquer modalidade ou instituição, inclusive da USP, exceto auxílios permanência (moradia e/ou alimentação);

c) não possuir punição por infração disciplinar registrada em qualquer época de seu curso atual.

2. VALOR E CARGA HORÁRIA: R$ 440,00 (quatrocentos e quarenta reais) mensais para o cumprimento da carga horária de 20 horas semanais.

3. VIGÊNCIA: de 2/5/2019 a 30/4/2020.

4. ATIVIDADES

Vide ANEXO I

5. INSCRIÇÃO

As inscrições serão presenciais, efetuadas mediante a entrega dos documentos abaixo relacionados no Escritório de Apoio à Pesquisa da FEARP (Bloco A, sala 40, telefone (16) 3315-4961) no período de 15 a 26/4/2019, diariamente, das 8h às 17h:

- Carta de motivação indicando como o candidato atende aos requisitos necessários;

- Histórico escolar de graduação com reprovações, se for o caso;

6. SELEÇÃO

A seleção será realizada pela análise dos documentos listados no item 4 e eventual entrevista presencial com o coordenador do projeto.

7. RESULTADO

O resultado será divulgado até às 18h do dia 29/4/2019 na página eletrônica da Pesquisa no sítio da FEARP (https://www.fearp.usp.br/pt-br/pesquisa).

ANEXO 1

Implementação de uma Rede Neural Multicamadas para a Classificação e Análise de Decisões Judiciais

Identificação da Proposta
Proposta apresentada para o edital de apoio a projetos que façam uso de sistemas digitais inteligentes (Portaria PRP nº 668, de 17 de outubro de 2018).
Proponente: Prof. Dr. Ildeberto Aparecido Rodello
Unidade: Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto (FEA-RP)
Modalidade: Universal

Resumo
A proposta é uma derivação do projeto de pesquisa intitulado "Mediação e Conciliação avaliadas empiricamente: jurimetria para proposição de ações eficientes", fomentado pelo Conselho Nacional de Justiça (CNJ) e desenvolvido pelo grupo de pesquisas "Applied Business Law", da FEA-RP. Usando um processo de extração, transformação e carga, foram obtidos os diários oficiais em formato *portable document format* que, depois de convertidos, foram carregados em um banco de dados NoSQL, criando uma base de dados reunindo as decisões judiciais publicadas no período de 2013 a 2017 dos estados: SP, RJ, PR, CE e PI. Um dos problemas encontrados durante a execução do projeto foram os diversos formatos de publicação dos diários oficiais, e também as suas diversas estruturas, o que levou a criação de uma base de dados mais bruta, mais genérica e menos detalhada. A proposta do sistema inteligente visa primeiramente realizar uma melhor classificação das decisões publicadas, categorizando-as nas classes de assuntos pertinentes, na natureza das partes envolvidas, no teor do objeto de discussão e nas decisões proferidas, dentre outros. Posteriormente, na medida que o sistema aprenda, pretende-se que ele seja importante para atualizações da base e também para a realização de análises mais apuradas, não somente em mediação em conciliação (primeira dimensão de treinamento do sistema), mas também em outros assuntos das áreas cível, criminal e trabalhista, por exemplo. Observa-se um grande potencial de contribuição do projeto, tanto para a gestão judiciária, quanto para orientação de advogados e pesquisadores da área.

1. Contextualização

Atualmente, a informação tem sido reconhecida como um ativo estratégico para qualquer organização e área de conhecimento. A gestão e a liderança eficazes envolvem, principalmente, a solução criativa de problemas, motivando os funcionários e garantindo que a organização, seja ela pública ou privada, atinja seus objetivos e metas.

Os dados obtidos de fontes internas e externas se constituem nesse ativo estratégico e são um desafio particularmente grande para os modernos sistemas de Tecnologia da Informação (TI), porque tornam esses sistemas muito lentos e não escaláveis o suficiente (LAUDON; LAUDON, 2015).

No que concerne a área de Direito, o acesso à informação é a ferramenta mais fundamental para garantir acesso qualificado à justiça. Apesar da previsão constitucional existente desde 1988 e que determina o acesso à informação como um direito individual dos cidadãos, o regramento infraconstitucional começou a regulamentar essas questões tardiamente.

Por um lado, aproveitando o poder dos recursos computacionais, os advogados podem prever com mais precisão como os eventos se desenrolam nos litígios. Assim, os advogados que adotarem a tomada de decisão baseada em dados terão uma clara vantagem sobre seus colegas que ainda se apegam a seus instrumentos desatualizados. (LETTIERI et al., 2018)

Não obstante, a demora na prestação jurisdicional e a ineficiência do judiciário são problemas amplamente conhecidos que não colaboram para a solução de conflitos sociais. A área jurídica se mostra invariavelmente restrita ao desenvolvimento de teorias jurídicas e processuais que, sem dúvida, são necessárias para o incremento do campo do direito, porém pouco auxiliam para a desejada entrega da pacificação social.

Dentro desse contexto, o projeto "Mediação e Conciliação avaliadas empiricamente: jurimetria para proposição de ações eficientes" é uma pesquisa empírica que está estudando formas de aperfeiçoar a entrega da prestação jurisdicional.

Trata-se de, inicialmente, coletar todos os dados referentes aos processos em que houve conciliação ou mediação nos estados estudados (CE, SP, PR, RJ e PI[1]). A tarefa não se demonstra simples e configura empreendimento necessário para um trabalho incontestável. Não se trata de definir uma amostragem e avaliar alguns processos. Trata-se de montar um banco de dados com todos os processos existentes - seja por meio do acesso a dados públicos dos tribunais ou mesmo por acesso às informações fornecidas pelo próprio Tribunal. Com base nesse banco de dados, estão sendo feitas inicialmente análises quantitativas sobre os processos, o que estão fornecendo informações extremamente relevantes para a tomada de decisões e propostas de ação, como o tempo médio de duração das demandas e os valores negociados, dentre outros.

É apenas com base nesses dados coletados que será possível chegar à conclusão de que tipo de ação é mais eficiente e eficaz para a melhoria dos processos internos que levem a decisões com melhores resultados (mais céleres e que alcancem, de fato, a pacificação social). Assim, após o estudo do que empiricamente tem acontecido nos Tribunais, serão determinadas e propostas as ações relevantes tendentes a aperfeiçoar a dinâmica das conciliações nos tribunais do país.

Espera-se que a utilização de um sistema digital inteligente possa prover mais facilidades para se alcançar os objetivos do projeto supracitado e mais acurácia para as análises dos dados.
Sob diversos aspectos, a redes neurais artificiais têm sido aplicadas com sucesso nos mais variados problemas. Dentre as principais áreas de aplicação de redes neurais artificiais, é possível citar o reconhecimento de padrões (BISHOP, 1995). No que se refere a área jurídica, é possível encontrar menções, ainda incipientes, do uso de redes neurais artificiais para auxiliar a dosimetria de penas (GRIMM, 2006) e no combate ao cyber crime (LOSSIO, 2017), dentre outros.

Nesse contexto, esta proposta visa a implementação de uma rede neural multicamadas, aplicada na área de Direito, mais especificamente ao projeto Mediação e Conciliação avaliadas empiricamente: jurimetria para proposição de ações eficientes", para realizar a classificação e posteriormente a análise de decisões judiciais publicadas nos diários oficiais.

2. A Base de Dados

A base de dados do projeto, onde a rede neural irá atuar, está sendo hospedada com recursos da Internuvem/USP e tem, no momento, a quantidade de documentos armazenados listada na Tabela 1.

Tabela 1 - Documentos (registros) armazenados na Base de Dados

Estado	Documentos
Ceará	3.132.367
Mato Grosso	3.868.862
Paraná	5.014.728
Piauí	509.450
Rio de Janeiro – Capital	14.410.301
Rio de Janeiro – Interior	13.101.916
São Paulo	90.345.365
Total	**130.382.959**

[1] Estados determinados pelo edital de convocação pública e de seleção convocação n. 02/2017 (EDITAL 02/2017, [s.d.]).

Como mencionado, os estados foram determinados pelo edital de convocação pública e de seleção convocação n. 02/2017 (EDITAL 02/2017, [s.d.]), do CNJ, compreendendo o período de 2013 a 2017 (5 anos). Porém, a intenção é criar a base de acesso gratuito, atualizada com todos os estados brasileiros. Dessa forma, é possível perceber o potencial que a base poderá proporcionar, não somente para pesquisas na área, como também para gestão da área jurídica.

3. Projeto do Sistema

Esta seção apresenta a visão geral inicial do funcionamento do sistema, especificando as entradas, o processamento e as saídas esperadas. Apresenta também os recursos em termos de linguagem de programação inicialmente planejados para sua execução.

3.1. Visão Geral Inicial do Funcionamento

Entradas

O sistema terá como entradas as sentenças das decisões judiciais publicadas e armazenadas na base de dados do projeto "Mediação e Conciliação avaliadas empiricamente: jurimetria para proposição de ações eficientes". Cada processo judicial publicado no Diário é um registro (na nomenclatura NoSQL, um documento), que deverá ser lido pelo sistema. A estrutura do documento se encontra em um formato JSON (*JavaScript Object Notation*), ou seja, em formato texto, completamente independente de linguagem.

O conteúdo dos diários foi segmentado programaticamente de modo que cada registro representasse apenas uma publicação. Em cada registro inserido constou-se, além do conteúdo da publicação, a data, os dados forenses relacionados e o número do processo, que serviu como número identificador, ou ID, no jargão tecnológico.

Assim, os dados coletados e transformados em arquivos texto, foram organizados de acordo com a estrutura da Figura 1.

Figura 1 - Exemplo da estrutura básica JSON para armazenamento dos da

{"_id":"5ac7cf311fbbd54d5e84f7f4","Dados_Fórum":"JUÍZO DE DIREITO DA 3ª VARA JUIZ(A) DE DIREITO LUCAS PEREIRA MORAES GARCIA ESCRIVÃ(O) JUDICIAL APARECIDO CAVALCANTE DE MEIRELES","Processo":"009037-43.2011.8.26.0268","Conteúdo":"Processo 0009037-43.2011.8.26.0268 (268.01.2011.009037) - Execução de Alimentos - Alimentos - M. S. B. - L. B. - Vistos. Com fundamento no art. 794, I, do Código de Processo Civil, julgo extinto o presente processo, diante da informação de fls. 26 dos presentes autos. Fixo honorários advocatícios em 100% do valor da tabela. Transitada em julgado e não havendo custas a serem recolhidas, expeça-se certidão e após, arquivem-se os autos com as cautelas de estilo. P. R. e Int. Itapecerica da Serra, 11 de dezembro de 2012. - ADV: LUIZ HENRIQUE DOS SANTOS (OAB 237245/SP)","Data":"20130108"}

No contexto das redes neurais artificiais, as entradas estão principalmente no campo "Conteúdo" da estrutura exibida na Figura 1 e se constituirão no conjunto de sinapses ou conexões de entrada que alimentarão o aprendizado do sistema.

Processamento

A partir da leitura do campo "Conteúdo de cada documento serão extraídas as palavras-chave (de acordo com a dimensão – inicialmente mediação e conciliação), onde o sistema deverá identificar, analisar e classificar, a priori:

- Os nomes e aprender como classificá-los em pessoa física, jurídica ou governo. Deverá também identificar o gênero: masculino, feminino. Por fim, identificar se o nome está em segredo de justiça.
- A respectiva classe do assunto que o processo se refere. As classes de assuntos deverão seguir as recomendações de classificação do CNJ.
- Os valores, caso haja valor envolvido no processo.

Como função de ativação da rede neural, pretende-se utilizar a função do tipo sigmoidal, principalmente por ser a mais utilizada em redes neurais artificiais.
Para a determinação das palavras-chave e do contexto do aprendizado da rede neural, os membros da equipe com conhecimento na área de direito terão papel fundamental.

Saídas

Como saída o sistema entregará uma nova estrutura JSON adaptável com as classificações e refinamentos, que poderá substituir a antiga estrutura do banco de dados, ou mesmo gerar um subconjunto (*data warehouse*) de acordo com o interesse. Classifica-se a estrutura como adaptável, pois, para cada contexto de processo, poderá haver diferenças de classe, assuntos, etc.

3.2. Recursos que serão utilizados

A arquitetura que se pretende implementar em termos de redes neurais é a multicamadas. Para isso, o sistema será desenvolvido na linguagem de programação Python[2] com o auxílio de algumas bibliotecas específicas, tais como a NLTK[3]. A NLTK é uma biblioteca Python específica para o processamento de linguagem natural, com recursos para classificação, *tokenização*, *stemming*, rotulação, análise e raciocínio semântico, principalmente.

O sistema de gerenciamento de banco de dados utilizado é o MongoDB[4], uma vez que toda a estrutura do projeto já está implementada. O MongoDB é um sistema gerenciador de banco de dados NoSQL, orientado a documentos.

4. Metodologia de Desenvolvimento

A metodologia de desenvolvimento do sistema, em termos da teoria de engenharia de software, será aquela classificada como ágil, mais notadamente a metodologia Scrum. Nesta seção apresentam-se as etapas previstas e o cronograma de execução do projeto como um todo.

4.1. Cronograma

Para o período de 1 ano (12 meses) estão planejadas as seguintes atividades, dispostas temporalmente no cronograma do Quadro 1

1. Levantamento do requisitos e entendimento detalhado das funcionalidades
2. Implementação do sistema. Os *sprints* (metodologia ágil), serão determinados a partir da etapa 1.
3. Testes, treinamento da rede e execução
4. Ajustes finos e geração da documentação
5. Redação do relatório final

Quadro 1 - Cronograma previsto

Ativ	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
1												
2												
3												
4												
5												

5. Resultados Esperados

[2] https://www.python.org/
[3] https://www.nltk.org/
[4] https://www.mongodb.com/

Espera-se a implementação de uma rede neural multicamadas capaz de fazer a classificação dos documentos armazenados no banco de dados de decisões judiciais publicadas nos Diários dos Tribunais Justiça e coletados para o projeto "Mediação e Conciliação avaliadas empiricamente: jurimetria para proposição de ações eficientes" (130.382.959 de documentos, conforme Tabela 1).

Inicialmente a classificação envolvendo partes, tema, classes e assuntos, dentre outros, será para os estados de SP, RJ, PI, CE e PR. Contudo, a intenção da base de dados é contemplar todos os estados brasileiros. Da mesma forma, a primeira dimensão de treinamento consistirá do assunto mediação e conciliação. Pretende-se, posteriormente, realizar treinamento da rede também para outras dimensões e temas, tais como: cível, trabalhista, judicialização da saúde, dentre outros.

Dessa forma, percebe-se a importância e a contribuição que o sistema poderá proporcionar, não somente para o projeto em tela, como também para futuras pesquisas.

Referências

BISHOP, C.M. (1995) "Neural Networks for Pattern Recognition", Oxford Un. Press, ISBN: 0198538642.

EDITAL 02/2017. "Edital de convocação pública e de seleção convocação". Disponível em: http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/780a1047c628a90da5a29928f0d4dc01.pdf. Acesso em: 15 maio. 2018.

GRIMM, C. (2006) "Dosimetria da pena utilizando redes neurais". Monografia. Curso de Direito. Universidade Federal do Paraná. Curitiba.

LAUDON, K.; LAUDON, J. (2015) "Sistemas de informações gerenciais". 11a. ed. [s.l.] Pearson/Prentice Hall.

LETTIERI, N. et al. (2018) "Ex Machina: Analytical platforms, Law and the Challenges of Computational Legal Science". Future Internet, v. 10, n. 5, 26 abr.

LOSSIO, C.J.B. (2017) "O anticrime nas redes sociais: os algoritmos e a rede neural artificial (RNA) em face do cybercrime". Disponível em https://jus.com.br/artigos/58938/o-anticrime-nas-redes-sociais. Acesso em 15/12/2018.