RICARDO VIEIRA DE CARVALHO FERNANDES
HENRIQUE ARAÚJO COSTA
ANGELO GAMBA PRATA DE CARVALHO

Coordenadores

TECNOLOGIA JURÍDICA E DIREITO DIGITAL

I CONGRESSO INTERNACIONAL DE DIREITO E TECNOLOGIA – 2017

Prefácio

Ana Frazão

Realização

Organização Executiva

Apoio

Belo Horizonte

2018

INTELIGÊNCIA ARTIFICIAL (IA) APLICADA AO DIREITO: COMO CONSTRUÍMOS A DRA. LUZIA, A PRIMEIRA PLATAFORMA DO BRASIL COM *MACHINE LEARNING* UTILIZADO SOBRE DECISÕES JUDICIAIS

RICARDO VIEIRA DE CARVALHO FERNANDES

DANILO BARROS MENDES

HUGO HONDA FERREIRA

ANDRÉ BERNARDES SOARES GUEDES

1.1 Introdução

Estatísticas judiciárias oficiais, promovidas pelo Conselho Nacional de Justiça – CNJ, confirmam o que milhões de jurisdicionados que aguardam uma tutela de direitos nos tribunais já sabem: a jurisdição estatal brasileira encontra-se em crise. Mais do que isso, é possível dizer que o Judiciário brasileiro está em situação muito difícil: a morosidade é absurda, a falta de efetividade flagrante e os procedimentos inadequados a atingir os fins almejados pelas partes.

Talvez isso seja herança de um velho Brasil, cujo aparelho estatal foi estruturado em uma burocracia extremamente ineficiente, complexa e corrupta, atolada em leis inexequíveis, incongruentes e disformes.[1]

[1] MARINONI, L. G.; BECKER, L. A. Influência das relações pessoais sobre a advocacia e o processo civil brasileiro. In: BECKER, L. A. (Org.). *Qual é o jogo do processo?* Porto Alegre: Sérgio Antônio Fabris Editor, 2012. p. 447-480.

Novas leis, sobretudo aquelas datadas do ano de 2015 – Novo CPC (Lei nº 13.105/15), Lei da Mediação (Lei nº 13.140/15) e Nova Arbitragem (Lei nº 13.129/15) – tentam melhorar esse cenário. Todavia, legislação pura, desacompanhada de uma mudança cultural, não será capaz de retirar do Judiciário a centralidade uníssona da resolução de conflitos.

Atualmente, são mais de 18 mil juízes e centenas de milhares de servidores no sistema judiciário brasileiro.[2] Em sua imensa maioria, esse grupo relevante de profissionais se dedicam com muito afinco a resolver com efetividade, celeridade e paixão seu ofício. Porém, acreditamos que o volume atual de processos impede uma solução sem o apoio de tecnologia.

Temos o Judiciário com o maior volume processual do mundo, sendo o TJSP o maior tribunal do mundo. Recordes estes que não nos orgulham. Mas o que está posto precisa ser pensado, debatido e, especialmente, melhorado ou resolvido.

A partir da estipulação do monopólio do Estado na prestação de justiça – uma das heranças do Estado Moderno –, a via judicial constituiu-se como a única forma de resolução de conflitos admitida no ordenamento jurídico. A autotutela foi afastada, mas o acesso sempre foi limitado.

Recentemente, eventos históricos brasileiros dão conta da abertura do acesso ao Judiciário. Após o período cinzento de repressão militar, o retorno à democracia registrou na Carta de 88 o valor essencial do livre acesso ao Judiciário. Nada (ou quase nada, *vide* o *non liquet*) poderia barrar as pessoas em sua pretensão de levar suas demandas à presença de um juiz.

O mundo era outro em 1988. Como lembra Werneck Vianna,[3] a própria Constituição foi parte do processo de transição do autoritarismo à democracia, e não uma conclusão dele. Isso impactou, nas décadas seguintes, no aumento exponencial do volume de processos entrantes no sistema judicial. À medida que o valor do acesso passa a ocupar o âmago das pessoas, o Brasil começa a levar seus problemas ao Judiciário.[4]

---

[2] O número total de colaboradores (força de trabalho) era, em 2016, de 442.345, sendo 18.011 magistrados; 279.013 servidores e 145.321 auxiliares, cf. CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 35. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017.

[3] VIANNA, L. W. *et al*. *A judicialização da política e das relações sociais no Brasil*. Rio de Janeiro: Revan, 1999. p. 38.

[4] Luiz Werneck Vianna assevera a mudança de posição do Judiciário ao longo dos anos, sobretudo após a Constituição de 1988; antes como um poder periférico, "encapsulado em

Em 1988, o volume de processos brasileiros era estimado em menos de 400 mil processos. Cerca de uma década depois, a estimativa deu conta de um salto, estimado entre 2 e 6 milhões de processos. Em 2003, primeiro ano do relatório *Justiça em números*, só na Justiça Estadual havia 3 milhões de processos em estoque e 3,7 milhões novos processos ingressaram naquele ano. O total de processos já ultrapassava 10 milhões em 2003. Cinco anos mais tarde, em 2008, já tínhamos 70,1 milhões de processos. Em 2014, ultrapassamos o marco de 100 milhões deles.[5]

A boa notícia é que a Justiça tem empatado o jogo: o número de processos entrantes é praticamente idêntico aos processos julgados nos últimos 3 anos. Segundo os sucessivos relatórios *Justiça em números*, respectivamente em relação a 2014, 2015 e 2016, de 21 a 28,8 milhões de casos novos chegaram ao Judiciário todos os anos (praticamente o mesmo número de julgamentos, especialmente em 2015 e 2016). A esse número são somados os respectivos 73 a 79,6 milhões de processos em estoque nos tribunais brasileiros, relativamente ao mesmo período.[6]

O resultado é o número de mais 100 milhões de processos em tramitação no Brasil nos últimos 3 anos. Em 2016 ingressaram 28,82 milhões de processos, foram baixados 28,87 milhões, enquanto o estoque estava em 78,97 milhões.[7]

Por opção ideológica, abrimos completamente a porta de entrada do Judiciário a partir de 1988. Nunca se pensou em atingir esse volume... Não pensamos na porta de saída. Agora precisamos resolver a saída, a morosidade, o estoque.

Além do grande volume há também uma assustadora lentidão no contexto judicial. O tempo médio de tramitação dos processos pendentes na Justiça Comum Estadual é de 5 anos e 4 meses na fase de

---

uma lógica com pretensões autopoiéticas inacessíveis aos leigos, distante das preocupações da agenda pública e dos atores sociais", passa a se mostrar como uma instituição central à democracia brasileira, inclusive "no que diz respeito à sua intervenção no âmbito social" (VIANNA, L. W. *et al*. *A judicialização da política e das relações sociais no Brasil*. Rio de Janeiro: Revan, 1999. p. 9).

5 A partir de 2003 as informações deste parágrafo são do relatório *Justiça em números* do CNJ (CNJ. *Justiça em números*. Disponível em: <http://www.cnj.jus.br/programas-e-acoes/pj-justica-em-numeros>. Acesso em: 10 out. 2017). Anteriormente as referências são esparsas e desencontradas, utilizamos especialmente as pesquisas de VIANNA, L. W. *et al*. *A judicialização da política e das relações sociais no Brasil*. Rio de Janeiro: Revan, 1999.

6 CNJ. *Justiça em números*. Disponível em: <http://www.cnj.jus.br/programas-e-acoes/pj-justica-em-numeros>. Acesso em: 10 out. 2017, restrito aos anos de 2014, 2015 e 2016.

7 CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 37. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017.

conhecimento no 1º Grau, 2 anos e 6 meses no 2º Grau e de 7 anos e 6 meses na Execução. Nos processos não pendentes é de 3 anos e 1 mês na fase de conhecimento no 1º Grau, 1 ano no 2º Grau e de 3 anos e 4 meses na Execução.[8] Tais dados confirmam que o cidadão comum, ao ver seus direitos subtraídos, vai esperar um período considerável de sua vida para receber "justiça": a prestação jurisdicional.

O decurso exagerado de tempo até a resolução da demanda distorce o sistema de Justiça, impondo ao credor o ônus do tempo pela demora judicial.[9] O culpado, o devedor sempre se beneficia da morosidade. O que também reflete negativamente na confiabilidade dos cidadãos no sistema judicial nacional.

Além da morosidade, um importante reflexo de um sistema de justiça exacerbado é o aumento de seu custo. A despesa total com o sistema judicial brasileiro no ano de 2003 foi de R$14 bilhões de reais, sendo que aproximadamente 86% desse gasto, ou R$12 bilhões, foi com pessoal.[10] Pasmem, somente a despesa com pessoal para a manutenção e expansão desse serviço, em 2016, foi de R$75,94 bilhões (ou 89,5%), de um total de gastos na ordem de R$84,8 bilhões.[11] O gasto com pessoal vem tendo um aumento significativo e constante ao longo do tempo.

Descendo nos dados de 2016, temos que foram gastos com tecnologia da informação R$2,4 bilhões.[12] Isto é, o gasto com tecnologia

---

8 CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 39. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017.

9 Nesse sentido, L. A. Becker fala do "bônus do tempo do processo", ao referir-se à influência do tempo sobre os custos das partes processuais, sobretudo no que se refere à antecipação ou não da tutela (BECKER, L. A. O dilema dos litigantes: processo civil e teoria dos jogos. In: BECKER, L. A. (Org.). *Qual é o jogo do processo?* Porto Alegre: Sérgio Antônio Fabris Editor, 2012. p. 76-77). Ver também BECKER, L. A. A erosão do sagrado processual. In: BECKER, L. A. (Org.). *Qual é o jogo do processo?* Porto Alegre: Sérgio Antônio Fabris Editor, 2012. p. 235-236. Conforme assevera Ivan Ribeiro "de fato, uma parte com melhores condições econômicas está em melhor posição para participar do caso, uma vez que tem mais facilidade de acesso à justiça, melhores advogados, mais recursos para suportar tanto as despesas do processo quanto o tempo para obter uma decisão, entre outras vantagens" (RIBEIRO, I. *Robin Hood versus King John*: como os juízes locais decidem casos no Brasil? p. 15. Disponível em: <http://www.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006_Profissional_01lugar_tema01.pdf>. Acesso em: 13 out. 2017).

10 CNJ. *Justiça em números 2003*. Brasília: CNJ, 2004. pg. 2-3; 5. Disponível em: <http://www.cnj.jus.br/images/stories/docs_cnj/relatorios/justica_numeros_2003.pdf>.

11 CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 36. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017.

12 CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 36. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017.

e inovação não chegou a 3% do total gasto pelo Judiciário Nacional. Quando lembramos do aumento crescente do volume de ingresso de processos e do tamanho do estoque, a conclusão direta é que essa conta não fecha e não fechará. A ampliação do serviço sem o uso de tecnologia inovadora escalável tende a manter o histórico de crescimento desarrazoado de gastos públicos com pessoal.

Os entraves apresentados para a existência de uma Justiça efetiva começaram a causar reação por parte do Poder Público. Uma das importantes vias para solução desse problema é o sistema multiportas incentivado pela citada legislação de 2015, em que se busca dividir com a sociedade a responsabilidade pela resolução de disputas, seja pela mediação extrajudicial e judicial – inclusive por meio de digital, o que incentivou a criação das *startups* que se propõem a realizar ODR (*on-line dispute resolution*) –,[13] seja pelo incentivo à arbitragem. Há ainda a reforma processual, com a adoção de mecanismos processuais de julgamento de massa, como o incidente de resolução de demandas repetitivas, entre diversos outros.

Outra via para solucionar o sistema de justiça é a utilização de inteligência artificial (e outras formas de inovação tecnológica, como a análise de redes complexas) para realização das tarefas repetitivas, seja no lado das partes e seus respectivos advogados, seja no lado do apoio ao Judiciário nas tarefas repetitivas, especialmente naquelas não decisórias, que atrasam a tramitação do processo, mas também no apoio a decisões.

No presente *paper* nos propusemos a explicar como foi a construção da primeira plataforma com inteligência artificial aplicada sobre decisões judiciais para apoiar o peticionamento em processos de massa, seus passos e desafios. Foi a percepção do enorme potencial da tecnologia de *machine learning* (incluindo *deep learning*) sobre as massas processuais que nos moveu à publicação.

Os Tribunais e os grandes litigantes – Poder Público, bancos, seguradoras, telecons, *e-commerce* etc. – são os que mais se beneficiarão da tecnologia.[14] Há, também, fortes aplicações ao contencioso

---

[13] Sobre o tema, *vide* Capítulo 3 da Parte I deste livro, *E-negotiation, e-mediation, and the expansion of online dispute resolution in Brazil*, de Ricardo Vieira de Carvalho Fernandes, Colin Rule, Taynara Tiemi Ono e Gabriel Estevam Botelho Cardoso.

[14] Acreditamos que o sistema judicial brasileiro é condescendente com o litígio de massa na medida em que não divide a conta com o litigante contumaz, seja com aumentos gradativos de custas e/ou com multas processuais por não resolver seus problemas internamente. Parte da conta dos custos judiciais deveria ser efetivamente repartida com

administrativo de massa, como as demandas do INSS, de Detrans etc. O impacto da tecnologia é muito potente para áreas jurídicas de massa, judiciais ou administrativas, públicas ou privadas, incluindo de empresas que buscam gerir melhor seu contingenciamento.

1.2 Os primeiros passos

O primeiro passo foi definir adequadamente o problema inicial a enfrentar. Para tanto, o recorte de nicho (restrição do objeto de aplicação do *machine learning* – ML) foi essencial. Não existe inteligência artificial genérica, conforme bem explica Manisha Salecha.[15] Escolhemos, então, o maior volume de ações judiciais brasileiras: as execuções fiscais, que acumulam em seus variados degraus de ineficiência[16] mais de 32 milhões de processos (acima de 30% do total de processos judiciais brasileiros).

Dados relevantes do domínio analisado (campo jurídico) são a base essencial necessária para a aplicação da inteligência artificial (IA). É necessário um grande empenho dos desenvolvedores da tecnologia, lado a lado com especialistas do domínio, para agregar volume e diversidade relevante de informações e as estruturar de forma coerente, tornando-as úteis para procedimentos de aprendizado de máquina.

Uma vez definido o nicho e os dados sobre os quais iríamos nos debruçar, passamos à fase de observação e organização dos dados. A extração dos dados é uma tarefa inicialmente custosa e realizada de acordo com a demanda da solução, pois os dados raramente estarão disponíveis em uma formatação ideal e utilizável pelo sistema.

Assim, um procedimento adequado é agregar os dados sob a estratégia *extract-transform-load* – ETL, que executa as tarefas de identificar informações relevantes na fonte de dados, extrair essa informação,

esses litigantes massivos. Esse é um debate essencial para a solução do problema de justiça sob os múltiplos processos. Contudo, *de lege lata*, no âmbito restrito deste artigo e com as observações do uso da tecnologia e do mercado, esses são os campos de maior utilização de IA.

15 Em texto sobre conceitos básicos de inteligência artificial estreita e inteligência artificial geral (SALECHA, M. Artificial narrow intelligence vs artificial general intelligence. *Analytics India Magazine*, Aug. 2017. Disponível em: <http://analyticsindiamag.com/artificial-narrow-intelligence-vs-artificial-general-intelligence/>. Acesso em: 10 out. 2017).

16 Entre os quais citamos: legislação defasada, cadastros ineficientes dos cobradores públicos, falhas na inscrição da CDA, bases de bens não compartilhadas entre entes, volume exponencial, ardilosas artimanhas de muitos dos grandes devedores, máquinas do Executivo e Judiciário despreparadas para o volume de processos etc.

personalizar e integrar essa informação com diversas fontes, limpá-las e transformá-las com base em regras predefinidas e carregá-las em algum tipo armazenamento estruturado.[17]

Todas as informações utilizadas para procedimentos de aprendizado de máquina pela plataforma Dra. Luzia precisaram passar por este processo para se tornarem úteis às operações de aprendizado realizadas posteriormente. Tivemos que transformar processos judiciais em dados passíveis de serem interpretados pela máquina.[18]

Devido à necessidade de um enorme volume de dados para aplicação de ML, tal tarefa se torna inviável de ser feita manualmente, de forma que a estratégia de ETL utilizada foi codificada em mecanismos automatizados e os dados extraídos massivamente foram estruturados e processados sob um controle de qualidade de forma eficaz, otimizando assim a execução de operações pela futura plataforma Dra. Luzia.

1.2.1 Sobre os dados

Informações utilizadas pela primeira etapa da plataforma Dra. Luzia foram extraídas a partir de dados públicos e disponíveis em fontes abertas. As origens que possibilitam maior captação de dados são páginas *web* de órgãos públicos e Tribunais em geral, que disponibilizam o acesso irrestrito a seus processos. Melhor que fossem dados abertos disponíveis, mas, em sua falta, as equipes extraem massivamente tais dados públicos – o que aumenta o custo do projeto, mas não o inviabiliza.

1.2.2 Extração (*extract*)

A técnica de obtenção de dados desestruturados disponíveis em uma página *web* é chamada raspagem de dados – *scraping* –, que é a análise programática de uma fonte de dados para poder acessar uma ou mais fontes *on-line* para procurar informações disponíveis, geralmente de forma aberta no código-fonte da página.

---

[17] SIMITSIS, A.; VASSILIADIS, P.; SELLIS, T. Optimizing ETL processes in data warehouses. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE 2005), 21st, 2005. *Proceedings*... Tokyo, 2005. p. 564-575.

[18] Gostaríamos de deixar registrado nosso agradecimento à equipe de mineração de dados/ETL do Professor Dr. Henrique Araújo Costa que participou conosco do trabalho multidisciplinar e complexo de definição e extração da camada de dados para aplicação de ML.

A raspagem de dados *web* é realizada a partir de um conjunto de técnicas capazes de obter automaticamente informações desejadas de um *site* em vez de copiá-las manualmente. O raspador busca um conjunto de informações para extrai-las e as agregar, transformando dados não estruturados em algo que possa ser salvo em bancos de dados de forma estruturada.[19] Nesta etapa já são utilizadas algumas técnicas de processamento de linguagem natural – PLN, ou NLP em inglês.

O código-fonte de uma página *web* é constituído por marcações que constituem cada elemento em sua interface visual. A partir da detecção destas marcações, estruturas, conteúdo de cada elemento *web*, pode-se obter os dados específicos desejados especificados pelo código de extração.

Foi feito um procedimento similar ao que ocorre com extração de *DJEs* ou andamentos de Tribunais, só que mais profundos e relacionados.

1.2.3 Transformação (*transform*)

A transformação dos dados é um pré-processamento que resulta no que virão a ser as informações utilizadas pelo sistema de aprendizado de máquina. Esta etapa foi otimizada em execução simultânea à extração, em que os dados extraídos das páginas *web* já recebiam por um processamento inicial para se adequarem à estruturação em formato JSON.[20]

Novamente são utilizadas técnicas de NLP de forma que erros sejam corrigidos imediatamente, evitando processar novamente a base de dados. Com esta conversão para o formato desejado em uma estrutura modelada à plataforma, cada processo é salvo como um arquivo no formato JSON e certo conjunto de arquivos ficou definido com um lote de dados que armazena os resultados da extração.

O formato primordial de armazenamento e estruturação escolhido foi o JSON devido à sua capacidade de interpretação cada vez mais universal e à compatibilidade com as diversas tecnologias de bancos de dados não relacionais – NoSQL –,[21] permitindo uma maior escalabilidade na utilização desses dados.

---

[19] A linguagem natural é espécie de dado não estruturado, mas o ETL é capaz de converter ocorrências de linguagem natural em representações mais formais e estruturadas, mais facilmente manipuláveis, facilitando processamentos computacionais.

[20] JSON. Disponível em: <http://www.json.org/json-pt.html>.

[21] NOSQL. Disponível em: <http://nosql-database.org/>.

1.2.4 Carregamento (*load*)

Com os dados já extraídos e pré-processados, é necessário armazená-los na arquitetura do sistema de forma segura, resiliente e com visão de escalabilidade. Desta forma foi escolhido um sistema de banco de dados NoSQL que, a partir de um código de análise sintática do formato JSON, recebe os dados dos lotes extraídos atualizando o sistema de forma consistente, de forma que processos que estão ativos tenham seu detalhamento atualizado de acordo com o avanço de seu andamento no Tribunal.

1.3 Estudo do objeto e programação estratégica

Uma vez que os dados foram carregados em uma base estruturada, foi possível estudá-los com uma visão mais macro do domínio, para que conclusões pudessem ser tiradas e estratégias fossem construídas. É importante que toda a equipe envolvida na construção da solução seja capaz de traçar estratégias e entender propostas de solução para o problema, tendo em vista uma equipe multidisciplinar (engenheiros de *software*, cientistas de dados, engenheiros de produção, cientistas da computação, juristas, pesquisadores etc.), como a equipe da Legal Labs, desenvolvedora da Dra. Luzia.

É essencial que exista um time de especialistas na área que externalize o conhecimento do domínio para o restante do grupo. É a relação mais profunda dos respectivos conhecimentos de área que permite e evolução da tecnologia. Conhecimentos multi e interdisciplinares transmitidos, retransmitidos em reflexo e debatidos são muito potentes para esse campo.

Um time que tem conhecimento e autoridade sobre o domínio pode tomar melhores decisões sobre como encarar a tecnologia e como solucionar problemas com menor esforço, não os tornando complexos sem necessidade.[22]

Para criar uma estratégia eficaz também foram considerados os conceitos de inteligência artificial estreita (*artificial narrow intelligence*, ou ANI em inglês) e inteligência artificial geral ou genérica (*artificial general intelligence*, ou AGI em inglês). A ANI é mais especialista no que

[22] LAANTI, M. Characteristics and principles of scaled agile. In: DINGSØYR, T. *et al*. (Ed.). *Agile methods*. Large-scale development, refactoring, testing, and estimation. International Conference on Agile Software Development. Nova York: Springer, 2014. p. 9-20.

faz, e sabe fazer apenas o que é treinada para fazer. A AGI, por outro lado, tem o foco mais amplo, no qual o objetivo é desenvolver uma tecnologia capaz de simular o pensamento humano, podendo tomar decisões quando necessário.[23]

Podemos dizer que uma inteligência geral seria aquela que tem como objetivo se capacitar de forma generalista e horizontal, e que realiza tarefas em diferentes áreas ou nichos de uma área. Uma inteligência estreita, por seu turno, é aquela que tenta se especializar ao máximo, de forma vertical, nas tarefas que lhe são concedidas. Por mais que se pense em implementar uma AGI, esse patamar de tomada de decisão ainda está a alguns anos de pesquisa à frente do nosso tempo.

Ademais, tendo em vista o objetivo de gerar o maior impacto com o menor esforço, como já mencionado, foi escolhida a área de execuções fiscais, a qual se pode verificar que em 2016 detinha 30,4 milhões de processos pendentes, com a maior taxa de congestionamento nacional, 91%.[24]

Desse contingente de processos, se destacaram processos que são mais mecânicos e, em sua maioria, necessitam de menos tomada de decisão para movê-los. Processos simples que ocupam o tempo de procuradores da Fazenda Nacional, procuradores de estado e procuradores de município, além de servidores públicos, como resultado observa-se elevado gasto de verba pública com procedimentos repetitivos.

O objetivo foi gerar inovação tecnológica capaz de realizar/apoiar as atividades repetitivas enquanto as atividades de média e alta complexidade passam a poder receber mais atenção dos profissionais especializados.

Primeiro, construímos um "fluxo" com as classes "estados do processo" que contemplaram 29 estados das execuções fiscais

---

[23] GOERTZEL, B.; PENNACHIN C. *Artificial general intelligence*. Nova York: Springer, 2007. v. 2, livro teórico sobre AGI e sua viabilidade no contexto atual, ainda indisponível na tecnologia atual. Essa obra também discorre sobre comparativos com ANI e sua área. *Vide*: SALECHA, M. Artificial narrow intelligence vs artificial general intelligence. *Analytics India Magazine*, Aug. 2017. Disponível em: <http://analyticsindiamag.com/artificial-narrow-intelligence-vs-artificial-general-intelligence/>. Acesso em: 10 out. 2017.

[24] CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 108. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017. Citamos: "Os processos de execução fiscal representam, aproximadamente, 38% do total de casos pendentes e 75% das execuções pendentes no Poder Judiciário. Os processos dessa classe apresentam alta taxa de congestionamento, 91%, ou seja, de cada cem processos de execução fiscal que tramitaram no ano de 2016, apenas 9 foram baixados" (CNJ. *Justiça em números* – Ano-base 2016. Brasília: CNJ, 2017. p. 111. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/904f097f215cf19a2838166729516b79.pdf>. Acesso em: 13 out. 2017).

brasileiras. Tratou-se, pois, de um problema de várias classes. Tais estados poderiam ou não indicar "nós" ou classes de petições. Essa dupla construção se baseou mais nos conhecimentos dos juristas da equipe, mas teve essencial colaboração da equipe de dados à medida em que a reverberação das primeiras escolhas sobre os dados promovia alterações no modelo estratégico.

Assim, com o objetivo de desafogar esses processos e dar maior impacto à tecnologia, foram escolhidas as classes mais recorrentes no fluxo, responsáveis por gerar um impacto em até 85% de processos realizados na carga semanal de processos dos funcionários públicos (número até então desconhecido).

À vista disso, podemos destacar duas atividades que ajudam a definir a melhor estratégia para atacar um problema de classificação em *machine learning*, dentro de um objeto de estudo considerável:

- Identificar classes mais recorrentes no domínio: visto que a construção da solução sobre essas classes gera, de pronto, maior impacto e maior satisfação dos usuários finais com menor esforço.
- Identificar as características mais relevantes para cada classe: isso é feito para que, uma vez coletadas essas características, seja possível separar amostras de cada classe apenas com base nestes termos relevantes.

Tal estratégia foi desenhada tendo em mente o conceito de ANI, já que tal inteligência tem a capacidade de tornar o resultado final mais assertivo. Muitas vezes executadas tais tarefas repetitivas, a máquina é capaz de apresentar melhor desempenho que um ser humano.[25]

Assim, a inteligência criada é um serviço, uma auxiliadora e não uma competidora, para os que trabalham com a área de execução fiscal, no lado da cobrança. Pode ser adaptada para utilização judicial relativa ao mesmo campo de execuções fiscais.

Portanto, o serviço prestado pela IA gera mais tempo para trabalhos nos quais realmente se necessita atenção e tomadas de decisões críticas, que podem ser executadas apenas por pessoas, que possuem conhecimento para tanto.

Nosso objetivo foi de deixar a máquina fazer a atividade repetitiva (orientada e supervisionada por seres humanos) e liberar o ser

[25] SALECHA, M. Artificial narrow intelligence vs artificial general intelligence. *Analytics India Magazine*, Aug. 2017. Disponível em: <http://analyticsindiamag.com/artificial-narrow-intelligence-vs-artificial-general-intelligence/>. Acesso em: 10 out. 2017.

humano (cujo custo é altíssimo, sobretudo aqueles mais especializados que passam por difícil seleção pública para ingresso e possuem os maiores salários) para as atividades estratégicas. Com isso, acreditamos que estamos colaborando para reduzir os gastos desnecessários do serviço público.

1.4 Pré-processamento

Como já foi dito na seção 1.3, dados são a matéria-prima de quaisquer algoritmos de ML,[26] e com um bom conjunto de dados existe maior chance de colher melhores resultados na resolução do problema posto.

Porém, não se pode afirmar que um bom conjunto de dados é o suficiente para atingir resultados satisfatórios, também é necessária a certeza de que o dado a ser utilizado está em escala e formato aceitáveis e que as características selecionadas estão coerentes com problema atacado. Para tanto, utilizam-se técnicas para transformar um conjunto de dados ou um conjunto de características desses dados: é o pré-processamento de dados.

Pré-processamento é o conjunto de técnicas que estão correlacionadas de formas complexas e que permeiam uma grande área de atuação,[27] na qual sua correta aplicação pode gerar impactos significantes na generalização de problemas supervisionados.[28]

Uma dessas áreas de atuação que se pode citar é o processamento de linguagem natural (NLP), que trata de transformar texto em características numéricas. Esse processo pode se tornar árduo e complexo conforme a quantidade de textos e de documentos se torna mais extensa.

Segundo Andrew Ng, um dos mais importantes pesquisadores do mundo de IA, incluindo NLP, "propor características é difícil, moroso e requer o conhecimento de especialistas no domínio. Quando se trata de aplicações de aprendizado, nós passamos muito tempo refinando características".[29]

---

[26] Isso considerando sistemas de IA baseados em dados. Embora mais limitados, estão no campo da IA os sistemas baseados em regras, cujas regras são previamente postas para apresentarem resultados esperados.

[27] TAN, P.-N. *et al*. *Introduction to data mining*. Pearson Education: India, 2006.

[28] KOTSIANTIS, S., KANELLOPOULOS, D.; PINTELAS, P. Data preprocessing for supervised leaning. *International Journal of Computer Science*, v. 1, n. 2, p. 111-117, 2006.

[29] NG, A. Machine learning and AI via brain simulations. *Stanford University Plenary Talk Slides*, 2011. Disponível em: <forum.stanford.edu/events/2011/2011slides/plenary/2011plenaryNg.pdf>. Acesso em: 14 out. 2017.

Portanto, a área de NLP torna-se um desafio, independentemente do domínio em que é tratada. Em específico, quando se tem em vista o domínio jurídico, pode-se afirmar que esse desafio se torna ainda mais complexo pelo intrincado vocabulário inerente à área. Assim, um time de especialistas para enfrentar tal empreitada se torna especialmente necessário.

Para a construção da Dra. Luzia foi utilizado um time de juristas especializados em dados, que foi capaz de explanar para um time multidisciplinar a complexidade de tais documentos e seu vocabulário. Essa tarefa foi executada seguindo um processo meticuloso e matemático. Apenas dessa forma foi possível transformar documentos jurídicos em algo que a máquina entendesse.

Dessa forma, foram utilizados os dados carregados (subseção II-D), que ao final somavam carca de 350 mil processos judiciais empilhados em formato JSON, cada qual com publicações, andamentos e demais dados. Escolheu-se como estratégia tomar o corpo textual de publicações como a fonte em que decisões são baseadas (publicações como sendo quaisquer despachos, decisões e certidões), isso por possuírem mais detalhes e informações sobre a vida do processo e suas etapas. Ao final, detínhamos cerca de 900 mil publicações, das quais foram analisadas mais de 400 mil publicações para classificação das classes mais recorrentes do domínio. Portanto, para tal quantidade de dados é necessária a utilização de técnicas conceituadas de pré-processamento e NLP, para criação de dados de treinamento e avaliação consistentes e compreensíveis para à máquina.

1.4.1 Pré-processamento textual

Pré-processamento textual é uma subárea de técnicas de tratamento de dados que pode ser subdividida em dois ramos: triagem de documentos e segmentação textual.[30] O primeiro é um conjunto de técnicas que trata de tarefas como codificação textual, identificação de idioma e identificação do texto, em que tais táticas têm o objetivo de transformar um documento ou arquivo digital em corpos textuais bem definidos. Essas técnicas foram aplicadas nos passos de ETL na seção 1.3. Entende-se o último ramo como um conjunto de técnicas que realizam a conversão de documentos bem definidos para palavras e frases que

[30] INDURKHYA, N.; DAMERAU, F. J. *Handbook of natural language processing*. Boca Raton: CRC Press, 2010. v. 2.

o compõe. Dadas tais definições, esta subseção será direcionada para a explanação de técnicas de segmentação textual.

Segmentação de palavras e frases ("tokenização", *tokenization* em inglês) e normalização textual são os principais componentes da segmentação textual, em que tais algoritmos transformam um documento textual (refere-se a documento como quaisquer corpos textuais, compostos apenas de uma frase ou vários parágrafos) em representações que facilitam o processamento de tal informação por máquinas e consequentemente algoritmos de aprendizagem.

1) *Tokenização: token* é o termo dado para a menor unidade de informação textual que uma máquina pode processar.[31] Dessa forma, tokenização é o processo de transformar um documento em *tokens*. Essa técnica tem complexidades diferentes para cada idioma a ser processado – idiomas com origem europeia apresentam menor complexidade, quando comparados com idiomas que utilizam sistemas logográficos (como chinês e japonês), com relação aos quais essa tarefa se torna muito complexa. Para realizar a tokenização de um documento é necessário encontrar os limites de cada palavra ou frase. Por esse motivo a complexidade se eleva quando falamos de logogramas. Visto isso, pode-se verificar um exemplo de tokenização em um documento na Tabela 1, em que cada palavra é isolada (ilustrado por aspas) e criada uma nova representação desse documento.

Tabela 1 – Exemplo de tokenização

Documento	Resultado
Os funcionários economizaram 2 horas.	"Os" "funcionários" "economizaram" "2" "horas" "."

2) *Normalização textual*: normalização textual é uma complementação à tokenização, em que se utilizam *tokens* para normalizar sua forma. Por meio de abordagens gramaticais é encontrado um formato comum para diferentes representações de *tokens* equivalentes. Um exemplo desta técnica está destacado na Tabela 2, em que se nota a transformação de palavras para o mais próximo de sua raiz gramatical.

[31] HARDENIYA, J. P. N. Natural language processing: Python and NLTK. Birmingham: Packt Publishing, 2016.

Para alcançar tal resultado, existem duas abordagens, "stemização" (do inglês, *steamming*) e "lematização" (do inglês, *leamming*). A primeira é, em termos gerais, o processo de retirada de sufixos dos termos para que seja computada apenas sua raiz. Já a segunda é a busca pelo lexema do *token* em questão – lexema é a unidade mínima de uma palavra com significado lexical. A diferença a se notar entre tais técnicas é a busca por uma representação comum entre *tokens* equivalentes, visto que a stemização busca a raiz do *token* sem se preocupar com o sentido léxico do termo, já a lematização procura manter o sentido léxico do termo, reduzindo flexões para sua forma léxica mínima.

A utilidade dessa técnica se mostra quando se deseja diminuir a quantidade futura de características para algoritmos de aprendizagem, analisar o vocabulário do domínio ou retirada de informação, dado que, com documentos normalizados pode-se buscar informações independentemente de flexões ou de quão rica é a morfologia do domínio.[32]

Tabela 2 – Exemplo de normalização

Abordagem	**Documento**	**Resultado**
Stemização	Os funcionários economizaram 2 horas.	"o" "funcion" "economiz" "2" "hor" "."
Lematização	Os funcionários economizaram 2 horas.	"o" "funcionário" "economiza" "2" "hora" "."

1.4.2 Processamento de linguagem natural

De forma geral, processamento de linguagem natural, ou NLP, objetiva entender como seres humanos entendem linguagem natural, para que então seja possível criar métodos nos quais máquinas possam utilizar deste conhecimento para entender e manipular textos, ou falas, em linguagem natural.[33]

[32] INDURKHYA, N.; DAMERAU, F. J. *Handbook of natural language processing*. Boca Raton: CRC Press, 2010. v. 2.

[33] CHOWDHURY, G. G. Natural language processing. *Annual Review of Information Science and Technology*, v. 37, n. 1, p. 51-89, 2003.

1) *Palavras de parada (stop words)*: palavras de parada são referentes a palavras que possuem pouco significado dentro do domínio. Tais palavras podem representar um ruído na amostra de documentos, quando representado em formato de *tokens*. Uma das técnicas de NLP é a retirada de *stop words*, já que elas não agregam em nada ao significado do texto.

Para manter o controle de *stop words*, normalmente, é construída uma lista de termos que se encaixam na definição (como um dicionário de *stop words* para o domínio específico). Porém, criar e manter essa lista não é uma tarefa simples ou clara, já que palavras de parada são intrinsecamente dependentes do domínio,[34] apesar de estudiosos tentarem criar listas genéricas para idiomas.

Porém, tal prática leva a equívocos, uma vez que uma lista genérica da língua inglesa possui o termo "the", que pode ser confundido pelo termo presente no nome próprio "The Who". Dessa forma, a retirada da palavra "the", no domínio de bandas, por exemplo, acarretaria a perda de significado. O mesmo se aplica quando o domínio em questão é o português jurídico brasileiro, no qual existem termos com valor para a área que, ao mesmo tempo, não possuem valor para a língua portuguesa em geral. Esse é mais um contexto em que é essencial a presença de um time especialista no domínio para distinguir tais palavras.

2) *Bag of words*: esse modelo de representação é o mais simples, já que tal método não se preocupa com a ordem das palavras ou gramática, e sim com a afirmação de se uma palavra pertence a uma lista de palavras ou não. Dessa forma, a representação do modelo de *bag of words* é similar à tokenização, se diferenciando pelo fato de criar uma lista contendo apenas *tokens* únicos. Então, ao final o modelo gera uma lista de termos únicos, ou vocabulário, do *corpus* (conjunto de documentos).

Normalmente, o modelo de *bag of words* é utilizado juntamente com técnicas de ponderação, em que pesos são dados aos *tokens* encontrados. A principal tática utilizada para ponderação é a frequência de termos (*term frequency* ou TF, em inglês), em que o algoritmo de ponderação é apenas a contagem de ocorrências de um *token* no *corpus*.[35] Outra técnica comum é o TF-IDF (*term frequency – inverse document frequency*) e N-Grams, em que, além da contagem de termos, existe a ponderação

[34] FOX, C. A stop list for general text. *ACM Sigir Forum*, v. 24, v.1-2, p. 19-21, 1989.

[35] HARDENIYA, J. P. N. Natural language processing: Python and NLTK. Birmingham: Packt Publishing, 2016.

com base em outros documentos. A técnica TF calcula um peso para cada termo com base em quão frequente ele é em um documento, mas não em vários documentos no *corpus*. Assim, se uma palavra apresenta tal propriedade, é provável que esse termo seja importante para o contexto daquele documento. Essa ponderação é calculada seguindo a Equação 1, em que se nota que o peso calculado para cada termo é maior quando um termo aparece várias vezes em menos documentos e se aproxima do mínimo quando um termo aparece na maioria dos documentos.[36]

$$tfidf_{(t,d)} = tf_{(t,d)} \times idf_{t} \quad (1),$$

em que,

$$idf_{t} = \frac{log\ (Total\ de\ documentos)}{Node\ doc.contendo\ termo}$$

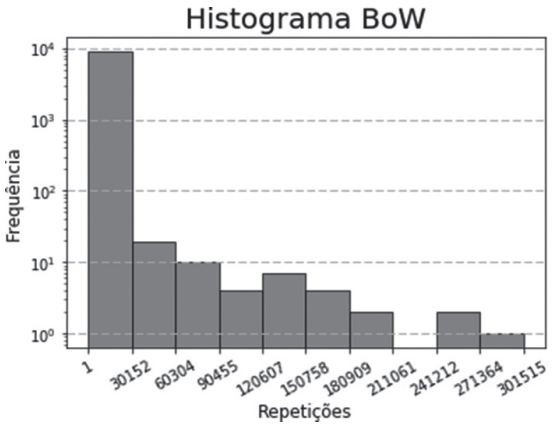
Tais técnicas são responsáveis por transformar uma lista de *tokens* em algo representativo e inteligível para a máquina, em que a representação textual de cada *token* é substituída por uma representação numérica. Assim, o princípio destes algoritmos é baseado em contar quantas ocorrências cada palavra possui, dentro de um conjunto de documentos, e substituir sua representação por seu número de ocorrências. Tal contagem pode levar a alguns problemas quando a lista de *tokens* não é tratada com cuidado, visto que a contagem de recorrência é baseada na exata semelhança dos *tokens*. Portanto, é necessária a execução correta da fase de pré-processamento antes de criar um modelo de representação *bag of words*.

Para o caso Dra. Luzia, ao analisar as quase 400 mil publicações, pré-processar os documentos e aplicar o *bag of words* nesse *corpus*, pudemos gerar histogramas (figuras 1 e 2) para visualização da quantidade de termos e suas respectivas repetições dentro do contexto proposto. A Figura 1 é referente a todo o *corpus* analisado, em que é possível notar que existem perto de dez mil termos raros (termos com poucas recorrências) e três termos que são muito frequentes (aparecem entre 241 mil e 315 mil vezes em todo o *corpus*). Exemplo de ambos pode ser visto na Tabela 3.

[36] MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. Scoring, term weighting and the vector space model. *Introduction to Information Retrieval*, v. 100, p. 2-4, 2008.

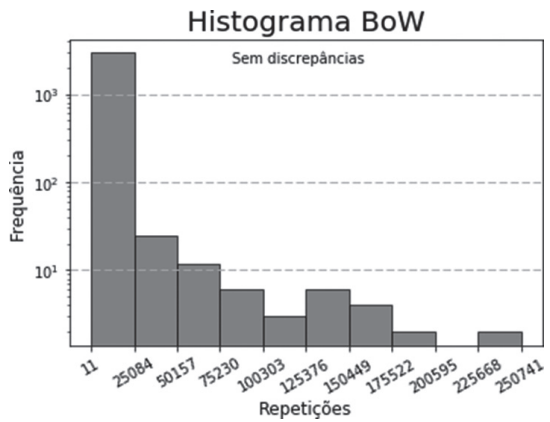
Percebe-se que os *tokens* raros provavelmente são referentes a nomes próprios ou termos não comuns ao domínio. Por outro lado, os termos mais frequentes são exacerbadamente comuns ao domínio e podem ser notados virtualmente em todo documento. À vista disso, ambos os casos são caracterizados como discrepâncias e se pode tomar como uma heurística: as duas situações apresentam *tokens* que não agregam valor ao significado do domínio.

Figura 1 – Histograma da técnica *bag of words*



Fonte: Autores.

Figura 2 – Histograma da técnica *bag of words* sem termos discrepantes



Fonte: Autores.

1.5 Amostragem dos dados

Amostragem é o nome dado para a técnica utilizada para selecionar, manipular e analisar uma parcela representativa do conjunto de dados, com o objetivo de identificar padrões e tendências. Tal técnica é dependente do contexto a ser aplicado, como escala e montante dos dados, assim como interações entre eles. Porém, aplicar tal técnica de forma eficaz pode resultar em vantagens como diminuição do custo de coleta dos dados e maior rapidez na medição de todo o conjunto.

Tabela 3 – Discrepâncias observadas

Tokens raros	*Tokens* frequentes
"remend","ariost", "éric", "baufak"	"var", "distrit", "feder", "fiscal"

No caso Dra. Luzia, em que a população de dados era toda de processos de execução fiscal de uma procuradoria fiscal, foi necessário criar estratégias para coletar eficientemente dados suficientes para o treinamento. Essa tarefa foi realizada em conjunto tanto pela equipe de especialistas no domínio jurídico, quanto pelos especialistas no domínio computacional, visto que o problema era de cunho universal, em que ambos poderiam contribuir para a eficiência.

Primeiramente, tomaram-se algumas poucas amostras das classes destacadas a partir dos 350 mil processos coletados, e então o time de especialistas jurídicos realizou, de forma manual, a tarefa de sublinhar termos importantes e recorrentes a cada classe. Em sequência foram criados dicionários de expressões regulares, baseados nos termos sublinhados, para coletar um grande número de amostras para cada classe, e então realizar a validação manual dos documentos coletados.

Para cada documento rejeitado no crivo, uma nova atualização no dicionário de expressões regulares foi realizada, gerando, assim, mais documentos para serem analisados e validados. Desse modo, tal sequência de tarefas foi realizada até que a equipe se sentisse confiante com o número de dados anotados coletados. Ao final, foram coletados 136.219 documentos anotados (manualmente e/ou com supervisão), que eram pertencentes a alguma classe das destacadas. Porém, para o treinamento também é necessário adicionar amostras de dados que não são pertencentes a nenhuma das classes, para que a máquina também aprenda o que ela não conhece. Consequentemente, o número final do conjunto de dados, com amostras positivas e negativas

para pertencentes de classes relevantes, foi calculado no montante de 226.219 documentos.

Com esse montante foi possível realizar uma separação "conservadora" do conjunto de dados, em que 60% dos dados foram utilizados para treinamento (D-train) e os 40% restantes para a avaliação e coleta de métricas dos modelos de aprendizagem gerados (D-test). Dessa forma, é possível garantir com mais confiança que os resultados gerados na avaliação não são tendenciosos por causa de uma amostra desbalanceada. Importante destacar que o D-test não pode conter exemplos conhecidos pela máquina, para que ela possa realmente ser "testada" em condições reais.

1.6 Aprendizado de máquina

As soluções utilizando aprendizado de máquina estão se popularizando nos mais diversos setores da sociedade contemporânea. Os sistemas de aprendizado de máquina são usados para identificar objetos em imagens, transcrever fala em texto, interpretar informações sobre notícias ou postagens, ofertar produtos de acordo com o interesse dos usuários e selecionar os resultados mais relevantes em uma pesquisa.[37] Assim percebe-se que as mesmas estratégias podem ser aplicadas ao universo jurídico de forma efetiva com o conhecimento técnico e jurídico adequado.

As técnicas convencionais de aprendizado de máquina estão determinadas por sua capacidade de processar dados: a partir de um sistema bem definido de reconhecimento de padrões, desenvolvido com engenharia de aprendizagem meticulosa e conhecimentos profundos do domínio analisado, busca-se uma adequada representação e estruturação dos dados que possa servir de entrada a um subsistema de aprendizagem que possa detectar ou classificar os padrões sobre os objetos jurídicos que lhe foram apresentados.

Muitas vezes esta aprendizagem de máquinas se dá de forma supervisionada, ou seja, cada tipo de dado já contém sua própria rotulação entre as informações utilizadas para treinar o algoritmo. Já sabemos do que se tratam certas instâncias e sabemos o que esperar como resultado a partir de determinada entrada, a classificação é previsível. Há diversas técnicas que resolvem casos de aprendizado

[37] MURPHY, K. P. *Machine learning*: a probabilistic perspective. Cambridge: MIT Press, 2012.

supervisionado: regressão linear, regressão logística, máquina de suporte vetorial (ou máquinas *kernel*), árvores de decisão, k-vizinhos mais próximos, *naive bayes* e redes neurais.[38] Basicamente, podemos obter inteligência a partir de conhecimentos predeterminados.

Outro tipo de abordagem é o aprendizado de máquina não supervisionado. Este não trabalha com dados pré-classificados, geralmente tenta agrupar os dados organicamente por suas semelhanças.

Existem também abordagens mais recentes, entre as quais está o aprendizado profundo, *deep learning*. Neste conjunto de técnicas se destacam as redes neurais profundas (DNN, do inglês *deep neural network*), que podem realizar o aprendizado tanto de forma não supervisionada como de forma supervisionada.[39] O grande diferencial desse tipo de estratégia é a capacidade de aprender representações intermediárias sem a necessidade de intervenção humana na modelagem das características que serão detectadas. Isto demonstra um potencial que ultrapassa todas as expectativas para esse tipo de algoritmo. Nesse sentido, a necessidade de conhecimento do domínio analisado é bem menor, pois a massa de dados analisados é capaz de ensinar à rede o que é mais importante na detecção dos padrões.

Esta aplicação se aproxima de forma natural da realidade cerebral, pois o aprendizado humano e animal é amplamente não supervisionado: descobrimos a estrutura do mundo observando-a, não sendo informados sempre acerca de cada característica dos objetos.[40]

No desenvolvimento da Dra. Luzia, houve algumas adaptações técnicas e estratégias de pré-processamento supervisionadas para alimentar uma rede profunda supervisionada principal, com uso de camadas de *deep learning* nas redes neurais.

1.6.1 *Naive bayes*

A ideia fundamental do *naive bayes* é que a partir de probabilidades estimadas para certo dado, pode-se decidir qual é o resultado mais provável. Desta forma, tal algoritmo é probabilístico e pode

---

[38] HONDA, H.; FACURE, M.; YAOHAO, P. Os três tipos de aprendizado de máquina. *LAMFO*, 27 jul. 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 10 out. 2017.

[39] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015.

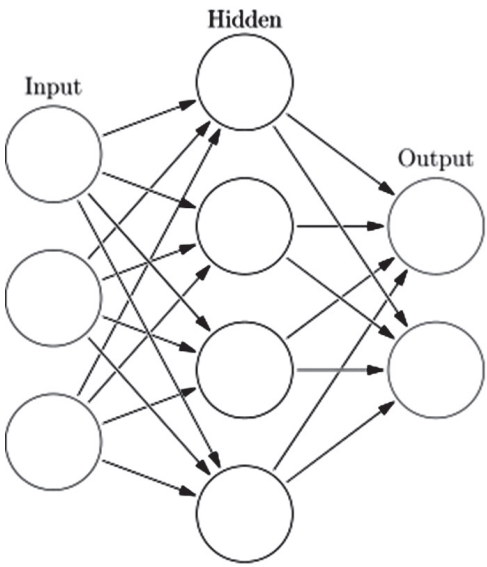
[40] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015.

ser definido pelo teorema de Bayes (Equação 2), em que se verifica a estimação de um resultado.

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

Desse modo, P (C X) é a probabilidade de uma amostra, ou documento no caso de classificação textual, ser de determinada classe C, dado que tal documento possui as características X.[41]

Figura 3 – Rede neural artificial simples pró-alimentada



Fonte: ARTIFICIAL neural network. *Wikipedia*. Disponível em: <https://en.wikipedia.org/wiki/Artificial_neural_network>.

Uma propriedade desse algoritmo é que os dados precisam ser linearmente separáveis, em que o cálculo de probabilidades leva em conta amostras próximas à avaliada. Visto isso, pode-se explicar

[41] ZHANG, H. The optimality of naive bayes. *AA*, v. 1, n. 2, p. 3, 2004.

a terminologia *naive* (ou ingênuo, em português) dada ao algoritmo, já que ele cria suposições que podem simplificar a realidade. Porém tal algoritmo pode se desempenhar com muito sucesso em alguns contextos, assimilando seus resultados a redes neurais, com muito mais complexidade.[42]

Todavia, para o contexto da Dra. Luzia, ao ser utilizado tal algoritmo os resultados demonstrados não foram promissores para o contexto de classificação de processos de execuções fiscais.

1.6.2 RNN – LSTM

Primeiramente, para entender RNN, deve-se citar que essa técnica é pertencente a um conjunto chamado *deep learning*. E tal conjunto possui características como a utilização de múltiplas camadas de unidades não lineares de processamento, que se comunicam entre si, aprendem em ambientes supervisionados e não supervisionados, assim como aprendem diferentes níveis de representação do dado analisado, em que pesos são treinados para cada comunicação entre unidades.[43]

RNN, ou rede neural recorrente, é uma classe de redes neurais artificial, que, por sua vez, são algoritmos e representações digitais que surgiram a partir de estudos biomiméticos do córtex cerebral e seus neurônios.[44] São representados receptores de informação (círculos vermelhos na Figura 3), com representação biológica equivalente aos 5 grandes receptores (olhos, ouvidos etc.), neurônios intermediários (círculos azuis), sinapses (setas) e neurônios terminais (círculos verdes), onde finaliza o processamento e o resultado é gerado.

Então, RNN é uma rede neural pró-alimentada (ou *feed foward neural network*, FFNN, em inglês), com a mudança de que a RNN tem noção de tempo e continuidade. Assim, a rede possui conexões entre camadas e através do tempo, em que alguns neurônios recebem informações tanto de camadas anteriores como deles mesmos de ações anteriores, possuindo certa memória. Dessa forma, uma rede neural recorrente é capaz de distinguir informações baseada não somente em

---

42 MITCHELL, T. M. *Machine learning*. [s.l.]: McGraw-Hill Science/Engineering/Math, 1997.

43 DENG, L.; YU, D. Deep learning: methods and applications. *Tech. Rep.*, May 2014. Disponível em: <https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/>. Acesso em: 10 out. 2017.

44 ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386, 1958.

dados instantâneos, mas também em dados passados.[45] Porém, um problema inerente de RNNs é a volatilidade de tais "memórias", que se perdem rapidamente a cada ciclo de aprendizagem.

Visto isso, foi criado o conceito de memória de longo/curto prazo (LSTM ou *long short-term memory* em inglês), que veio para tentar mitigar o problema de memórias voláteis. Esta técnica é inspirada majoritariamente em circuitos digitais e não tanto na biologia humana, em que cada neurônio ganhou portas, ou portões, que têm o papel de esquecer dados, receber dados ou emitir dados.[46] Com isso, a nova rede neural (RNN – LSTM) é capaz de aprender dependências de longo prazo, de modo que problemas como a máquina entender o contexto de uma frase ao chegar ao seu final se tornam mais simples.

A leitura de publicações extensas não se torna mais um problema para a máquina, já que ela consegue criar relacionamentos de longo termo entre os dados, ou seja, ela consegue identificar que o ato de um juiz citar uma lei no começo de uma sentença pode estar interligado à decisão descrita ao final do texto.

1.6.3 CNN – redes neurais convolucionais

Atualmente, com a evolução do *hardware* e o avanço da computação paralela, a utilização de redes neurais está cada vez mais proeminente. Não obstante, redes neurais convencionais ainda necessitam de um nível de processamento muito elevado, mesmo com uma baixa quantidade de camadas. Isso se dá porque essas redes, chamadas de "totalmente conectadas", contêm um número quadrático de conexões entre as camadas, o que torna o treinamento muito lento.[47]

Com o avanço nas pesquisas sobre o córtex visual animal, foi constatado que a organização dos neurônios e suas correlações são mais esparsas que as do cérebro humano,[48] em que pequenas regiões do espaço visual eram responsáveis pelos estímulos que acionavam neurônios individualmente. Essas regiões são conhecidas como campos receptivos.

---

[45] ELMAN, J. L. Finding structure in time. *Cognitive Science*, v. 14, n. 2, p. 179-211, 1990.

[46] HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.

[47] HAYKIN, S. *Neural networks*: a comprehensive foundation. 2. ed. New Jersey: Pearson Education, Prentice Hall, 1999.

[48] HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, v. 195, n. 1, p. 215-243, 1968.

As redes neurais convolucionais (detalhadas a seguir) são baseadas nesses princípios para minimizar o custo computacional necessário para seu treinamento. Quando comparadas com as redes neurais multicamadas tradicionais, o ganho de desempenho é muito significativo, e é ainda mais evidenciado pelo surgimento de algoritmos eficientes para computação de tais redes usando GPUs.

Algumas características importantes diferenciam esse tipo de rede neural dos demais tipos. Por conta da correlação mais esparsa dos neurônios, características espacialmente afastadas terão ligações bem fracas, se não nulas, entre si. Isso torna a rede mais capaz de generalizar características afastadas individualmente, porém também dificulta o uso desse tipo de rede quando existem relações importantes entre características que não estão próximas espacialmente. Imagens em geral muitas vezes obedecem a essa restrição, pois características visuais só agregam valor semântico quando espacialmente aproximadas, porém outros campos têm sido explorados com grande sucesso, como a classificação de texto.[49]

No contexto da construção da Dra. Luzia, camadas de redes neurais convolucionais foram usadas com o intuito de aumentar a discriminação da rede, encontrando as características locais entre as palavras dos textos das publicações.

1.6.4 RNN – Multicamadas

Para a implementação do cérebro da Dra. Luzia, foi utilizada uma rede neural recorrente, que possui diferentes camadas, cada qual com uma habilidade e função específica. Empilhou-se cada camada, para que a saída de uma camada complexa fosse a entrada da próxima, resultando ao final apenas as saídas necessárias para classificar um processo entre as classes iniciais.

A título de exemplo, ao realizar o treinamento de tal abordagem sobre 180 mil publicações (base para algumas classes), chegou-se ao número de parâmetros igual a 19.356.703, sendo que cada unidade é referente a parâmetros treináveis, ou seja, sinapses da rede neural final.

[49] ZHANG, X.; LECUN, Y. Text understanding from scratch. *CoRR*, 2015. Disponível em: <http://arxiv.org/abs/1502.01710>. Acesso em: 10 out. 2017.

1.6.5 Estratégias de treinamento

Após a decisão de quais algoritmos de treinamento utilizar e com quais dados alimentar a rede neural, é necessário realizar o treinamento. Tarefa que pode ser realizada de diversas formas, obtendo diferentes resultados a cada vez. Tais técnicas podem diferir de forma que variem a distribuição amostral dos dados, as formas de pré-processamento dos dados (seja o conjunto inteiro ou apenas algumas características), e os conjuntos de características. Cada aplicação de estratégia também necessita de um estudo prévio do algoritmo a ser aplicado à nova abordagem, pois alguns algoritmos de *machine learning* têm necessidades especiais, quando se refere a dados. Sendo que alguns não aceitam características não normalizadas ou com formatos diferentes do padrão estabelecido.

Além de procedimentos realizados sobre dados, ainda existe a possibilidade de refinamento de hiperparâmetros, em que para cada algoritmo ocorrerá variações no conjunto destes parâmetros livres, que podem ser definidos *a priori* à aprendizagem. Tal refinamento pode otimizar o aprendizado, tanto em sua assertividade, como em sua efetividade, esta verificada no tempo em que o treinamento demora para ser realizado. Para o âmbito de redes neurais artificiais, podem-se citar alguns parâmetros:

- *Learning rate*, que é a taxa de aprendizado do modelo: quão rápido a rede deve aprender novos conceitos e esquecer antigos.
- *Batch size*, que é um número inteiro que determina quantos documentos simultâneos irão ser treinados.
- *Optimizer*, que é o algoritmo utilizado para otimizar a rede neural.
- *Epochs*, que é quantas vezes a rede a ser treinada irá se treinar sobre os conjuntos inteiros de dados.
- *Activation*, que é referente à função de ativação dos neurônios: tal função tem o papel de decidir se uma informação deve ser passada à frente ou não.

Devido a tantos algoritmos e parâmetros livres, além da quantidade de opções de pré-processamento de dados, tal procedimento pode se tornar árduo e moroso. Existem algumas técnicas de automação para este processo de refinamento de parâmetros e definição de estratégias de treinamento. Porém, segundo Bergstra e Bengio,[50] para

[50] BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, p. 281-305, Feb 2012.

o procedimento de refinamento em redes neurais, é mais vantajoso realizar tal tarefa a partir de buscas randômicas a executar métodos automáticos de busca e refinamento.

1.7 Avaliação dos modelos

Para realização da avaliação dos modelos gerados foram utilizados os 40% dos dados anotados e separados previamente (seção 1.5). Visto que, com esses dados em mãos, é possível realizar a predição para cada modelo gerado e coletar uma matriz de confusão, como ilustrado na Tabela 4 para um caso de classificação binária.

Tabela 4 – Modelo de matriz de confusão

		Predição	
Real		V	N
	V	Verdadeiros positivos (VP)	Falsos negativos (FN)
	N	Falsos verdadeiros (FV)	Verdadeiros negativos (VN)

Tal tabela é uma técnica utilizada para gerar métricas de acurácia, dados os valores preditos e os valores reais, assim utiliza-se a Equação 3 para calcular a acurácia do modelo.[51]

$$acurácia = \frac{VP + VN}{VP + VN + FV + FN}$$

Apenas derivando mais valores da matriz de confusão podemos calcular outras métricas, como: precisão, *recall*, medida F1, curva ROC (*receiving operating characteris- tic*, em inglês) e área embaixo da curva ROC. Cada qual com suas vantagens e finalidades.

Para o contexto da Dra. Luzia, foi coletada a métrica de acurácia, pois ela supriu as necessidades impostas. Assim, podem-se visualizar na Tabela 5 os resultados finais coletados utilizando uma rede neural recorrente multicamadas.

[51] FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861-874, 2006.

Tabela 5 – Resultados finais utilizando RNN – Multicamadas

Classe	Acurácia
1	99,42%
2	99,68%
3	99,42%
...	...

1.8 Conclusão

Com este trabalho, em cerca de 10 meses conseguimos idealizar, estruturar, desenvolver, validar e colocar em produção a primeira plataforma do país capaz de tomar decisões jurídicas e escolher qual petição oferecer para dar andamento à massa de execuções.

Dada uma carga real de execuções fiscais com decisões judiciais diversas, ela escolhe, sem contato humano, entre as 8 petições-modelo inseridas em seu banco, qual o "estado do processo" atual e para qual "estado" a decisão judicial deslocou o processo. Isso é feito pelo *classificador*, que decide qual petição oferecer a partir das redes neurais explicadas neste *paper*. A partir daí o *processador* é capaz de buscar no banco as informações requeridas pelo magistrado em sua decisão/despacho, como endereços, veículos, imóveis. Por fim, essa informação é passada para o *gerador de petição*, que preenche as petições e as gera. Ao final, a equipe do órgão precisa observar o resultado e corrigir eventuais equívocos. Há ainda um *dashboard*, cujas informações capturadas pelo ML são mostradas ao usuário para tomada de decisões estratégicas.

A Dra. Luzia não faz petições complexas, nem invade a competência humana de tomar decisões. Ela só faz o que foi treinada para fazer. E sempre precisará de um advogado-usuário para verificar sua atuação. É *só o começo: um breve sopro do que há por vir*. É *o início do uso da tecnologia de machine learning, incluindo deep learning, sobre massa de dados jurídicos. O processo era físico, depois virou digital, inauguramos a próxima fase: inteligência artificial sobre contextos jurídicos*.

Quando o Judiciário começar a utilizar IA para processar tantos dados, o estoque de processos de massa tende a diminuir vertiginosamente e, com ele, o custo do sistema de justiça nacional. O mesmo deve ocorrer com grandes escritórios (privados ou públicos, como as advocacias públicas) e departamentos jurídicos de empresas.

Acreditamos, inclusive, que a tecnologia pode reduzir o Custo Brasil.

São inúmeras possíveis utilizações que não se confundem com "ser julgado pela máquina" ou "substituir o advogado" e que têm extrema potência para auxiliar os operadores jurídicos e ajudar a solver nosso problema de justiça.

Informação bibliográfica deste texto, conforme a NBR 6023:2002 da Associação Brasileira de Normas Técnicas (ABNT):

FERNANDES, Ricardo Vieira de Carvalho et al. Inteligência artificial (IA) aplicada ao direito: como construímos a Dra. Luzia, a primeira plataforma do Brasil com *machine learning* utilizado sobre decisões judiciais. In: FERNANDES, Ricardo Vieira de Carvalho; COSTA, Henrique Araújo; CARVALHO, Angelo Gamba Prata de (Coord.). *Tecnologia jurídica e direito digital:* I Congresso Internacional de Direito e Tecnologia - 2017. Belo Horizonte: Fórum, 2018. p. 39-67. ISBN 978-85-450-0453-0.