

Aplicando Deep Learning e NLP em Documentos Jurídicos

Pedro Vitor Quinta de Castro
pvcastro@gmail.com

Data Lawyer
Deep Learning Brasil

27 de Fevereiro de 2020



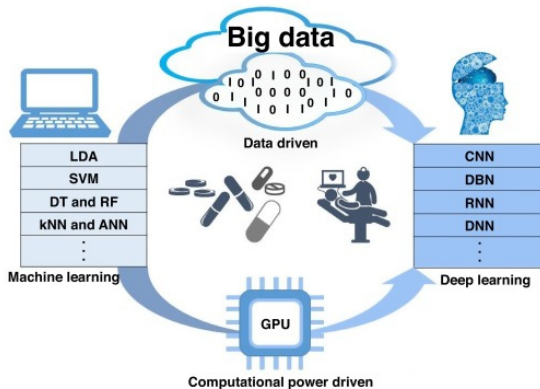
DEEP LEARNING
BRASIL

- Formado em Engenharia de Computação pela **UFG** em 2005
- Trabalho com Desenvolvimento de Software há mais de 15 anos
- Trabalho com Machine Learning há 5 anos
- Mestre em Processamento de Linguagem Natural pela **UFG**
- Pesquisador do grupo **Deep Learning Brasil**
- *CPO e Data Scientist* na **Data Lawyer**

Agenda

- 1 Inteligência Artificial e Direito
- 2 Jurimetria
- 3 Extração de Informações
- 4 Concepção e Arquitetura
- 5 Data Lawyer Insights

Inteligência Artificial



Machine Learning + Big Data + GPU = Deep Learning

Big Data



4,321,490,345

Internet Users in the world



1,710,695,288

Total number of Websites



157,889,204,397

Emails sent [today](#)



4,119,701,055

Google searches [today](#)



3,922,136

Blog posts written [today](#)



462,820,318

Tweets sent [today](#)



4,305,230,182

Videos viewed [today](#)
on YouTube



50,218,737

Photos uploaded [today](#)
on Instagram

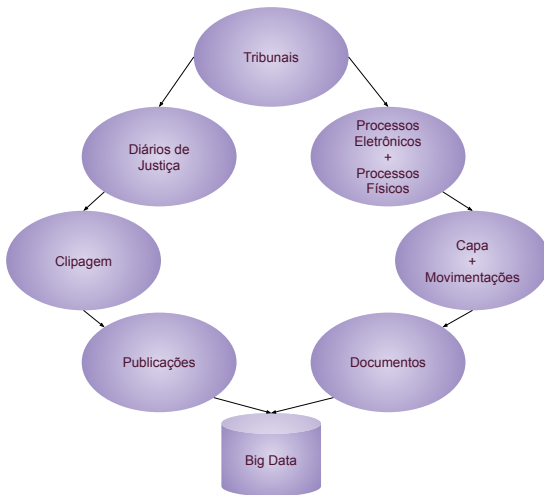


84,586,053

Tumblr posts [today](#)

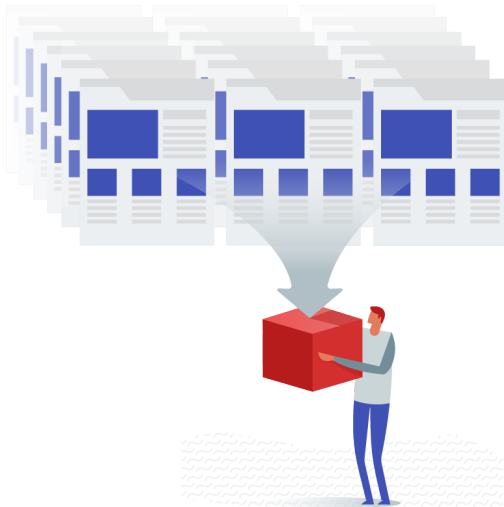
<https://www.internetlivestats.com/>

Big Data no Direito



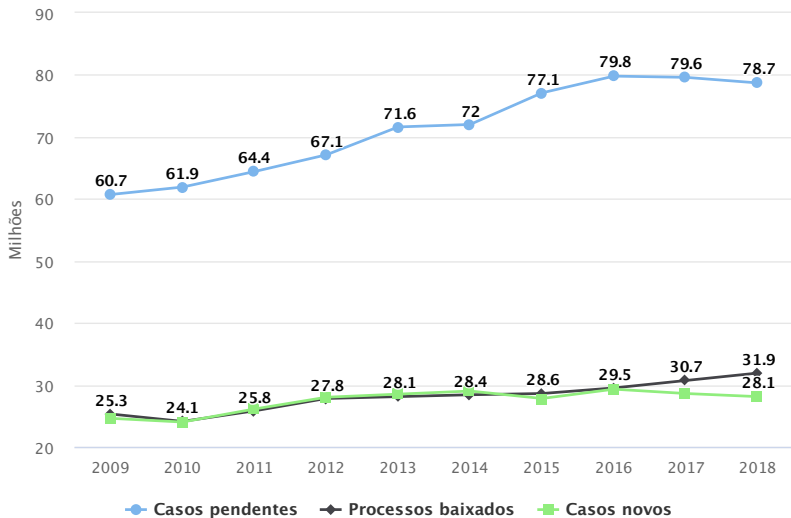
Dados produzidos na justiça

Como obter os dados?



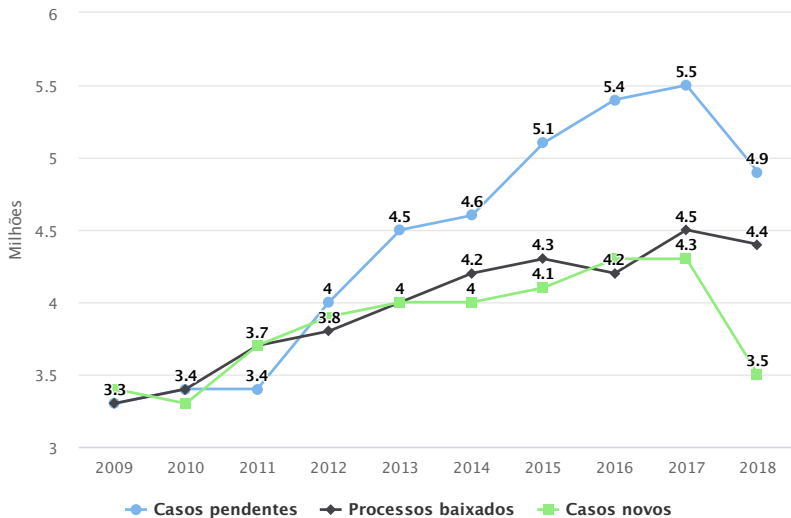
Crawling + Scraping

Oportunidade na Justiça Brasileira



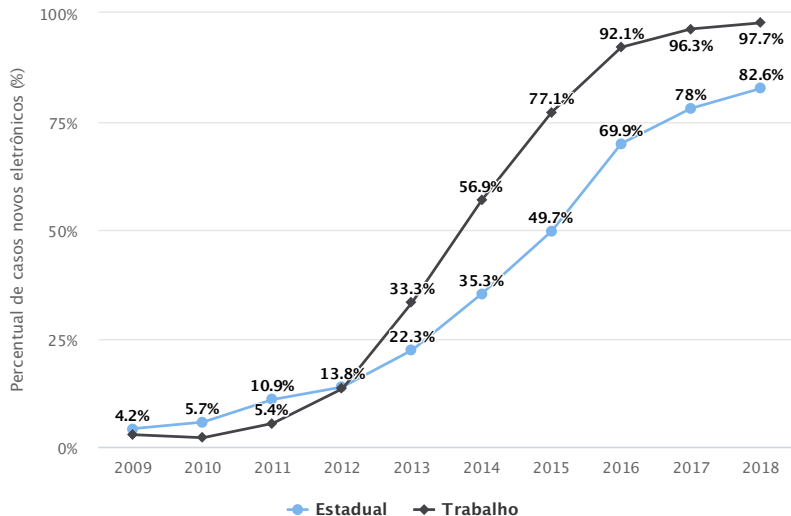
Highcharts.com

Na Justiça do Trabalho



Highcharts.com

Por que a Justiça Trabalhista?



Highcharts.com

Big Data de Processos

- 11.7 milhões de processos eletrônicos trabalhistas de 2014 até o presente
- 74 milhões de documentos indexados
- 300 milhões de anexos armazenados
- 40 a 50 mil processos atualizados diariamente a partir do processamento de todos os diários de justiça trabalhista
- Legal Research



Conceito Aplicação de modelos estatísticos em informações jurídicas

Objetivo Mapear comportamento da justiça por meio de análises estatísticas

- Qual o valor médio de condenação por juiz ou por foro?
- Como variam valores de acordo em relação aos valores originais de causa?
- Quais os dispositivos legais mais utilizados para fundamentação de juízes e advogados para cada tipo de pedido?



EFEITOS DA REFORMA

Na contramão dos danos morais, pedidos de hora extra sobem em São Paulo

15 de junho de 2019, 7h03

[Imprimir](#) [Enviar](#)

[Por Fernando Martinez](#)

A Reforma Trabalhista (Lei 13.467/2017) teve como um de seus grandes objetivos diminuir o número total de ações judiciais no setor. E, logo após ser implementada, em novembro de 2017, teve o efeito desejado. Alguns advogados alegavam que o fenômeno era efeito do represamento de ações, em uma espera por jurisprudência. Mas a fixação de sucumbência em caso de derrota do trabalhador parece ter um efeito que veio para ficar.

Números da **Data Lawyer**, empresa que coleta dados das cortes trabalhistas, mostram o efeito: um ano antes da reforma entrar em vigor, de 1º de outubro de 2016 até 31 de outubro de 2017, foram 779.560 casos novos somando 2ª e 15ª Regiões (São Paulo e interior paulista, respectivamente). Um ano depois, entre 1º de novembro de 2017 e 30 de novembro de 2018, foram 555.028 casos.

Os pedidos de dano moral seguem a tendência de queda: 107.035 casos antes da reforma e 47.003 após, analisando os mesmos tribunais nos mesmos períodos.

Porém, utilizando os mesmos parâmetros para checar os pedidos de horas extras, os dados mostram que estes aumentaram. Foram 100.644 casos antes da reforma e 146.031 depois, um aumento de quase 50%.

A **Conjur** conversou com advogados da área para ouvir teses sobre a dinâmica das ações de horas extras.

<https://www.conjur.com.br/2019-jun-15/contramao-danos-morais-pedido-s-hora-extra-sobem-sp>

Exemplos



O reconhecimento de vínculo empregatício nos serviços digitais de transporte

Published on December 20, 2019



Caio Santos
CEO na Data Lawyer

5 articles

✓ Following

<https://www.linkedin.com/pulse/o-reconhecimento-de-vínculo-empregatício-nos-serviços-caio-santos/>

Extração de Informações

- Como extrair estas informações?
 - Aquisição de dados não-estruturados
 - *Crawling + Scraping*: HTML + Documentos (PDF, RTF, etc...)
 - Estruturação via modelos de **Processamento de Linguagem Natural**
 - Extração de Entidades
 - Classificação de Texto
 - Clusterização

Extração de Entidades

- **Conceito**

- **Identificação e Classificação** automática de nomes próprios em texto
- REN - Reconhecimento de Entidades Nomeadas
- NER - Named Entity Recognition
- NERC - Named Entity Recognition and Classification
- Pré-requisito para **Extração de Relações** and **Vínculo de Entidades**

- **Exemplo**

Jair Messias Bolsonaro foi eleito presidente do **Brasil** pelo **Partido Social Liberal** nas eleições de **2018**

- **Domínio Geral**

- Pessoa, Organização, Local, Valor, Tempo, Outros

- **Domínio Legal**

- Função: *reclamante, reclamado, advogado, juiz, testemunha*, etc.
- Fundamento: "*artigo 795 da CLT*"ou "*artigo 1º da Lei n.º 6.858, de 24 de Novembro de 1980*"
- Tribunal, Vara
- Valores de Acordo, Causa, Condenação e Custas
- Local, Organização, Pessoa

Classificação de Texto

- **Classificação de Funções:** como determinar se uma função é de um juiz?

Des	MINISTRO
Des (a) .	Magistrados
Desembargador	Magistrados da 1ª Câmara
Desembargador Federal	Magistrados da 9ª Turma
Desembargador Relator	Mediador
Desembargador do Trabalho	Min .
Desembargador do Trabalho - Relator	Min . Rel .
Desembargadora Redatora	Min . Relator
Desembargadora Relatora	Ministro relator
Desembargadora-Relatora	REDATOR
Desembargadoras	REDATOR DESIGNADO
Desembargadores do Trabalho	RELATOR (A)
Des ^a	RELATOR DESIGNADO
JUIZ	REVISORA
JUIZ (A) DO TRABALHO	Redator
JUIZ SENTENCIANTE	Redator Designado
JUIZ TITULAR	Redator designado Ministro
JUIZA SENTENCIANTE	Redatora
Juiz (a) Titular de Vara do Trabalho	Rel
Juiz (a) do Trabalho	Rel . Des . Fed .
Juiz (a) do Trabalho ,	Rel . Ministro

Clusterização

- **Detecção de Valores Anômalos**

Vou cadastrar com qualquer valor...



Dados cadastrados...

Detalhes do processo ATOrd 0011539-72.2014.5.01.0016

Órgão julgador:

16ª Vara do Trabalho do Rio de Janeiro

Distribuído:

30/10/2014 16:17

Autuado:

30/10/2014 16:17

Valor da causa:

R\$ 40.000,00

☒ Processo com justiça gratuita deferida

Assunto(s):

☆ Anotação na CTPS

Anotação / Baixa / Retificação

Aviso Prévio

Décimo Terceiro Salário Proporcional

Férias Proporcionais

Grupo Econômico

Liberação / Entrega das Guias

Multa de 40% do FGTS

Multa do Artigo 467 da CLT

Multa do Artigo 477 da CLT

Saldo de Salário

Polo ativo

Reclamante:

R. C. A.

 [REDACTED] (ADVOGADO)

 [REDACTED] (ADVOGADO)

III - D E C I S Ã O

Isso posto, decide o Juízo da 16ª Vara do Trabalho do Rio de Janeiro julgar IMPROCEDENTES os pedidos da presente demanda, conforme exposto na fundamentação, que a esta passa a integrar.

Atribuí-se à causa o valor de **R\$ 38.000,00**, com custas no importe de R\$ 760,00, pelo reclamante, em face do disposto no artigo 852-A, da CLT.

Cumprimento em 08 dias. Intime-se.

Do que, para constar, editou-se a presente ata, que vai devidamente assinada.

ERICO SANTOS DA GAMA E SOUZA

Juiz do Trabalho

...por causa de 2 mil reais?



E nesse caso?

XPTO XPTO XPTO XPTO invoca a tutela jurisdicional desta Vara para propor reclamação trabalhista em face de **COMPANHIA DE SANEAMENTO BÁSICO DO ESTADO DE SÃO PAULO - SABESP**. Alega ter sido admitido em 19.07.1988, para exercer a função de ajudante de almoxarifado, promovido a motorista, com contrato suspenso por aposentadoria por invalidez. Além dos requerimentos de praxe, pleiteia as seguintes verbas e os títulos: (i) integração de adicional por tempo de serviço nas horas extras; (ii) diferença pela aplicação do divisor 200 até dezembro de 2012; (iii) doença profissional e danos morais e (iv) honorários advocatícios. Dá à causa o valor de **R\$ 50.000,00**. Junta documentos.

Diferença de mais de *5 bilhões*

Detalhes do processo ATOrd 1001053-81.2015.5.02.0716

Órgão julgador:

16ª Vara do Trabalho de São Paulo - Zona Sul

Distribuído:

19/06/2015 18:57

Autuado:

19/06/2015 18:57

Valor da causa:

R\$ 5.005.005.000,00

☒ Processo com justiça gratuita deferida

Assunto(s):

☆ Horas Extras

Adicional de Antiguidade

Divisor

Doença Ocupacional

Reflexos

Polo ativo

Reclamante:

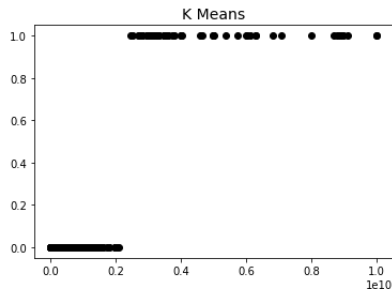
J. A. B. B.

  (ADVOGADO)

Clusterização por KMeans

- **71** processos
- **+320 bilhões** de valores de causa

```
In [83]: km = KMeans(n_clusters=2)
km.fit(X)
km.predict(X)
labels = km.labels_
#Plotting
plt.plot(X, labels, 'o', color='black');
plt.title("K Means", fontsize=14);
```



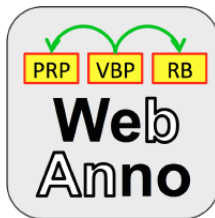
```
In [84]: print(Counter(km.labels_))

Counter({0: 9802674, 1: 67})
```

- Processo de treino do modelo de NER:
 - Criação de um dataset para treino supervisionado
 - Modelos de representações de palavras
 - Arquitetura do modelo
- Implementação da plataforma de Jurimetria
 - Avaliação de diferentes arquiteturas de armazenamento dos dados
 - Plataformas de BI

Criação do Dataset

- Seleção de documentos públicos de processos trabalhistas
 - Atas de Audiência, Sentenças e Acórdãos
 - Representatividade por Tipo de Documento, Tribunal e Ano
- Anotação das classes de entidades por um especialista
- Revisão para garantia dos critérios de anotação de cada classe



Exemplos de Anotações no WebAnno

10	Às 10h41min, aberta a audiência, foram, de ordem da Exma. Juíza do Trabalho, apreoadas as partes.
11	Ausente o exequente. Presente o(a) advogado(a), Dr(a). Priscila dos Santos, OAB nº 76251/RS.
12	Presente o sócio do executado, Sr(a). Nilvo Krummenauer, acompanhado(a) do(a) advogado(a), Dr(a). SILVIO LUIZ TASSINARI, OAB nº 32640/RS.
13	CONCILIAÇÃO:

	<u>FUNDAMENTO</u>	<u>PEDIDO</u>
	nos termos do art. 795 da CLT. DANO MORAL E MATERIAL. PRAZO PRESCRICIONAL. ACIDENTE DO TRABALHO OCORRIDO NA VIGÊNCIA DO CÓDIGO CIVIL DE 1916. AÇÃO PROPOSTA APÓS A EMENDA CONSTITUCIONAL Nº 45/04. REGRAS DE TRANSIÇÃO. Considerando que, ao início da vigência do Código Civil de 2002, havia decorrido mais da metade do lapso temporal da prescrição vintenária, observada a regra de transição consagrada no	
	<u>FUNDAMENTO</u>	
	art. 2028 do Código Civil de 2002, aplica-se à hipótese o prazo prescricional de vinte anos previsto no Código Civil de 1916, em respeito ao princípio da segurança	

279	<u>DECISAO</u>	<u>PEDIDO</u>
	Face ao exposto, acolho o pedido da inicial e condeno o Réu ao pagamento de honorários advocatícios calculados em 15% sobre o valor arbitrado à condenação, o qual, em observância ao valor fixado na inicial e aos pedidos acolhidos, delimita-se em R\$ 30.000,00 (trinta mil reais).	
	<u>VALOR_CONDENACAO</u>	
280	<u>FUNDAMENTO</u>	
	Assim sendo, conquanto as custas processuais, na forma do art. 789 da CLT, incidem em 2% sobre o valor arbitrado à condenação, fixo as custas, devidas pelo Réu, no montante de R\$ 600,00 (seiscentos reais).	
	<u>VALOR_CUSTAS</u>	

Documentos Anotados

- 144 documentos anotados + valores

Categoria	Anotação Final		Anotação Final + Valores	
	Entidades	Tokens	Entidades	Tokens
FUNCAO	1.435	2.881	1.689	3.327
FUNDAMENTO	2.639	19.547	3.347	24.475
LOCAL	271	965	392	1.284
ORGANIZACAO	1.773	5.690	2.317	7.791
PESSOA	1.580	5.566	2.064	7.158
TRIBUNAL	530	1.927	630	2.530
VALOR_ACORDO	23	68	337	1.010
VALOR_CAUSA	4	11	236	708
VALOR_CONDENACAO	30	84	457	1.358
VALOR_CUSTAS	64	176	829	2.465
VARA	209	1.427	238	1.595
Total	8.558	38.342	12.536	53.701

Seleção de Modelos de Representações de Palavras

- Representação em nível de caracteres por rede CNN
- Avaliação de Modelos Pré-Treinados:
 - Vetores Estáticos Pré-Treinados
 - FastText, GloVe, Wang2Vec, Word2Vec
 - Modelos de Linguagem
 - Bert, Albert, ELMo, Flair, spaCy
 - Pré-treino de vetores e modelos em acervo jurídico de mais de 6 bilhões de palavras

Seleção da Arquitetura do Modelo

- Arquitetura de Deep Learning para Classificação Sequencial: BiLSTM-CRF
- Avaliação dos Frameworks:
 - Spacy*
 - Flair
 - pytorch-transformers*
 - AllenNLP
- Modelo Final: AllenNLP com ELMo Jurídico + GloVe / Word2Vec Jurídico: **92% *F-Score***

AllenNLP

Exemplo de Extração com o Modelo

DESISTÊNCIA .

Requer o reclamante desistência da ação ajuizada , com consequente extinção do feito sem resolução do mérito , nos termos do [art. 485 , VIII do NCPC](#) .

Pois bem .

Em sendo assim , conforme [art. 200 do CPC](#) , homologo o pedido de desistência , declarando extinto o processo sem resolução do mérito , com base no [art. 485 , VIII do CPC](#) .

Retire-se o processo de pauta e arquivem-se os autos .

BENEFÍCIO DA JUSTIÇA GRATUITA .

Deiro ao reclamante tal benesse , pelo preenchimento dos requisitos do [art. 790 , § 3º da CLT](#) e da [Lei n.º 1.060 / 50](#) .

III.

DISPOSITIVO .

Isto posto e por tudo o mais que conste nos autos da Reclamação Trabalhista movida por [MAURO DE OLIVEIRA MIRANDA](#) e m face de [EMPRESA BRASILEIRA DE CORREIOS E TELEGRAFOS - ECT](#) , homologo o pedido de desistência , nos termos do [art. 200 do CPC](#) , e EXTINGO O PROCESSO SEM RESOLUÇÃO DO MÉRITO.Custas pelo recorrente no importe de [R \\$ 3.351,58](#) , calculadas sobre [R \\$ 167.578,85](#) , dispensadas face à gratuidade de justiça , valendo ressaltar que só na hipótese d e não comparecimento é que são devidas , e não na hipótese de desistência .

Assinatura

[MANAUS](#) , 20 de Março de 2019

[ALEXANDRO SILVA ALVES](#)

Juiz (a) do Trabalho Substituto

Assinado eletronicamente .

A Certificação Digital pertence a :

[[ALEXANDRO SILVA ALVES](#)]

[https : // pje.trt11.jus.br / primeirograu / Processo / ConsultaDocumento / listView.seam](https://pje.trt11.jus.br/primeirograu/Processo/ConsultaDocumento/listView.seam)

Imprimir

Exemplo de Extração com o Modelo

PODER JUDICIÁRIO FEDERAL

JUSTIÇA DO TRABALHO

TRIBUNAL REGIONAL DO TRABALHO DA 7ª REGIÃO

ÚNICA VARA DO TRABALHO DE SÃO GONÇALO DO AMARANTE

Ata de audiência do processo nº 0001639-23.2017.5.07.0039

RECLAMANTE: ANTONIO RAIMUNDO DA SILVA

RECLAMADO: RIO VERDE ENGENHARIA E CONSTRUÇÕES LTDA

Aos 18 dias do mês outubro de 2018, nesta cidade de São Gonçalo do Amarante, às 11h41min, estando aberta a sessão da Única Vara do Trabalho de São Gonçalo do Amarante, na Sala de Audiências, situada na Rua Santos Dumont, nº 28, Centro, sob a direção da Exmo(a). Juíza do Trabalho, Dr(a). MANUELA DE ALBUQUERQUE VIANA, foram por ordem da Sr(a). Juíza apregoados os litigantes do processo identificado em epígrafe.

Presente o reclamante, acompanhado do(a) advogado(a), Dr(a). ROBERTO PONTES SOARES, OAB nº 33984/CE.

Presente o preposto do(a) reclamado(s) RIO VERDE ENGENHARIA E CONSTRUÇÕES LTDA, Sr(a). RODRIGO DE CASTRO NAVARRO, acompanhado(a) do(a) advogado(a), Dr(a). NAHIANA DOS SANTOS ARAUJO, OAB nº 27277/CE. Requerido e deferido o prazo de 5 dias para juntada de substabelecimento e carta de preposição.

Presente o preposto do(a) reclamado(s) POSCO ENGENHARIA E CONSTRUÇÃO DO BRASIL LTDA, Sr(a). CARLOS HENRIQUE DE CASTRO FREITAS SOARES, acompanhado(a) do(a) advogado(a), Dr(a). GUSTAVO PUCCI SCHAUmann FILHO, OAB nº 30271/CE.

Presente o preposto do(a) reclamado(s) CSP - COMPANHIA SIDERURGICA DO PECÉM, Sr(a). LEANDRO ESTEVAN SOBREIRA, acompanhado(a) do(a) advogado(a), Dr(a). LUCAS SHALON CARDOSO DE ABREU, OAB nº 27408/CE.

Pela ordem, requereu a advogada da 1ª reclamada, perícia médica para constatar se a doença do reclamante possui nexos causais com o trabalho. Pela juíza foi deferida a perícia médica que deverá se realizar por carta precatória, uma vez que o reclamante mora em Piripiri no Piauí. Prazo de 10 dias para as partes, em querendo, apresentarem quesitos e assistentes técnicos.

Determina, ainda a juíza, que a Secretaria efetue a baixa na CTPS do reclamante em 15/05/2017, em face da incontrovérsia.

Considerando que o reclamante mora no Piauí, fica dispensada a sua presença em audiência; caso haja necessidade de se ouvir o reclamante, este deverá ser ouvido por carta precatória no momento oportuno.

Autos sobrestados até a realização da perícia médica.

Encerrada a audiência às 12h01min.

MANUELA DE ALBUQUERQUE VIANA

Juíza do Trabalho

A autenticidade do presente documento pode ser confirmada através de consulta ao site <https://pje.trt7.jus.br/documentos>, digitando a numeração que se encontra ao final do presente documento, abaixo do código de barras.

Ata redigida por p/A ROBERTA CORREA MARTINS CARVALHO, Secretário(a) de Audiência.

Arquitetura da Plataforma

- ✗ Stack de Arquitetura Lambda com Hadoop
 - Gargalo nos *Crawlers*
 - Manutenção cara e complexa
- ✓ **Apache Solr** para indexação de documentos
- ✓ **Elasticsearch** para indexação de dados estruturados



Visualização dos Dados

✗ Power BI

- Dificuldade de lidar com filtros
- Limitação com magnitude dos dados

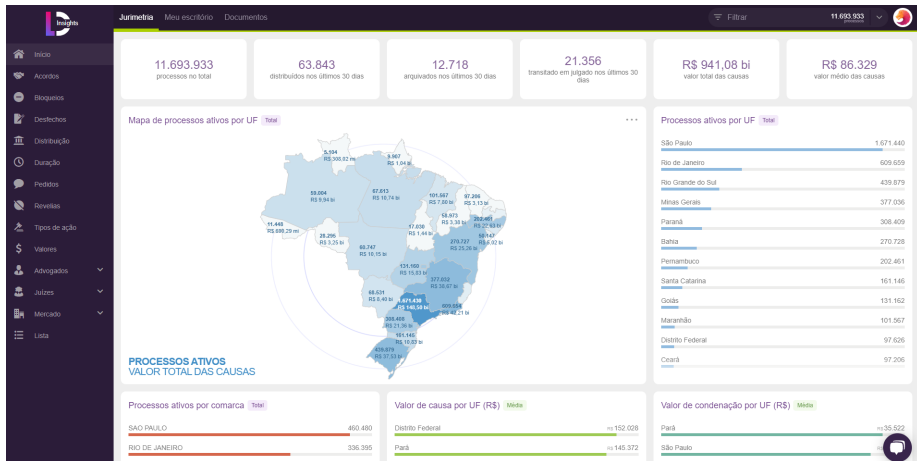
✓ Backend com **Elasticsearch**

✓ **Highcharts**

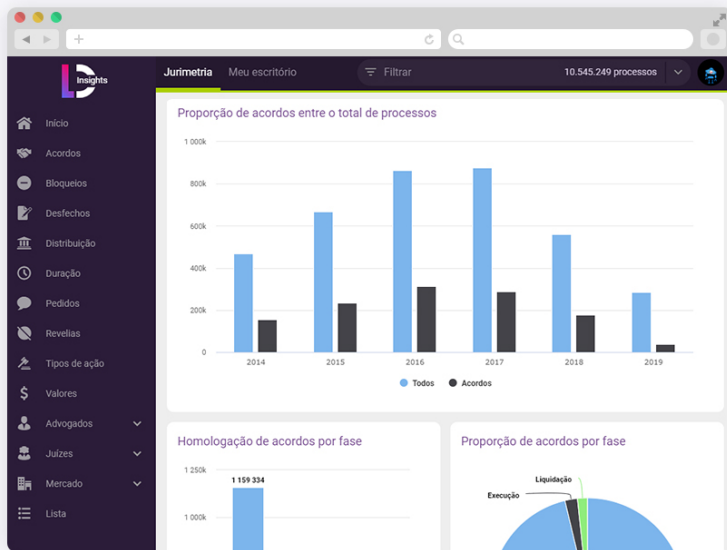


HIGHCHARTS

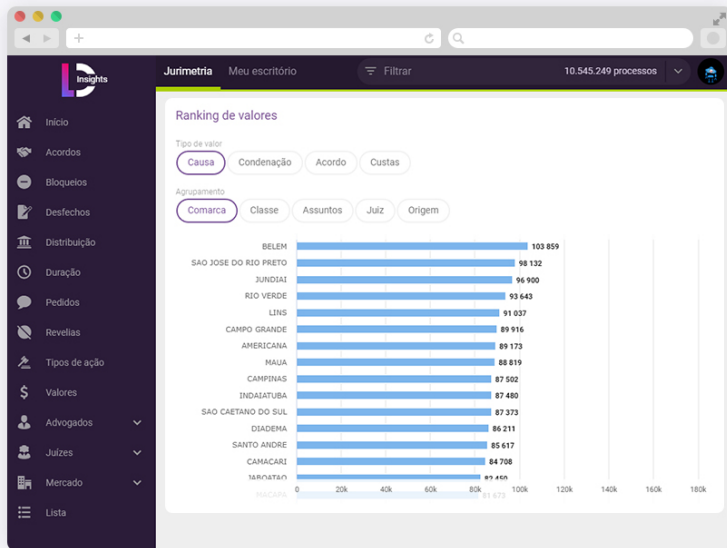
Data Lawyer Insights - Dashboard



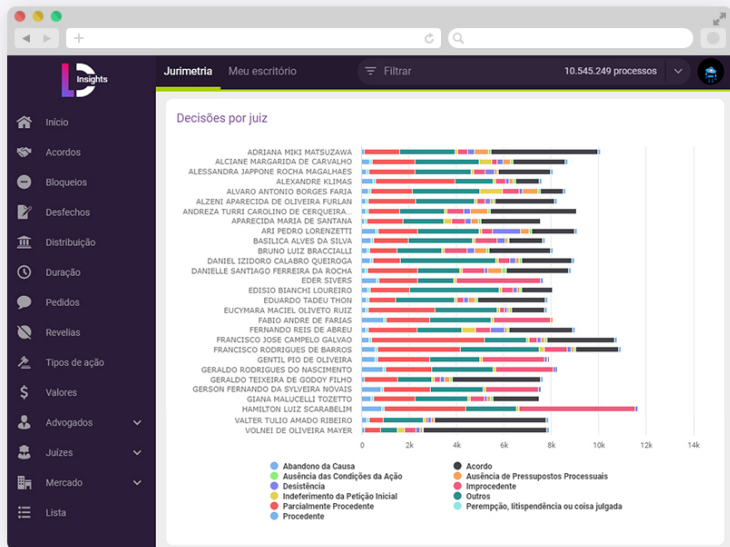
Data Lawyer Insights - Acordos



Data Lawyer Insights - Valores



Data Lawyer Insights - Juizes



Trabalhos Paralelos e Futuros

- Sugestão de tarefas para advogados a partir de publicações
- Identificação de Pedidos
- Detecção de Argumentos

Sugestão de Tarefas e Prazos para Advogados

SANTO ANDRÉ

Cível

1ª Vara Cível

JUIZO DE DIREITO DA 1ª VARA CÍVEL

JUIZ(A) DE DIREITO EDMUNDO LELLIS FILHO

ESCRIVÃO(O) JUDICIAL TAISA ESPER DE FREITAS ABUD

EDITAL DE INTIMAÇÃO DE ADVOGADOS

RELAÇÃO Nº 0482/2018

Processo 1016005-43.2018.8.26.0554 - Procedimento Comum - Auxílio-Acidente (Art. 86) - Luiz Carlos da Silva - INSTITUTO

NACIONAL DO SEGURO SOCIAL - INSS - Pedro Rodrigues Sanches - Defiro a gratuidade processual ao autor.

Anot-se. Quanto

à imediata implantação do benefício, inviável o acolhimento de pedido de concessão de tutela antecipada diante da inexistência

de probabilidade do direito. Apenas após realização da perícia é que se poderá decidir pela concessão ou não do benefício.

Não bastasse isso, o benefício previdenciário possui natureza alimentar, logo impossível ser repetido, razão pela qual não há o

necessário caráter de reversibilidade a viabilizar a tutela provisória perseguida. Evidente a necessidade de produção de prova

pericial médica e, ademais, por não vislumbrar prejuízo algum às partes, por questão de economia e celeridade processuais,

determino a realização antecipada da prova técnica, para tanto, nomeio perito o DRA. MARIA GONÇALVES

Fixo

seus honorários nos termos da portaria em vigor. Assim, nos termos do art. 321 do CPC, determino que o

procurador (a) da parte exequente emende a petição inicial, no prazo de 15 (quinze) dias.

Os interessados poderão acompanhar a perícia e providenciarão eventuais críticas. Exame do acidentado para o dia 20 de

agosto de 2018, às

- SP, intimando-

se o autor pessoal

la 65/68 6º andar Santo André

E. e providenciará o

SUGESTÃO DE ATIVIDADE



Emendar a petição inicial (art. 321 do CPC)

Prazo: 15 dias



+ Compromisso



[illegible]

- pvcastro@gmail.com

Agradecimentos



Dúvidas?