

## GT 2 "Análise de Dados: técnicas e ferramentas voltadas à boa governança de dados" - 19/08/2024, 15h às 18h

Sala 2 – SEGJUD, Bloco A – Sala A4.132

Link para Sala Virtual: [meet.google.com/zqs-ckao-bxy](https://meet.google.com/zqs-ckao-bxy)

Mediadora: Camila Ribeiro Rocha Torres (SEPJD)

Título do Trabalho	Autor(a)(es)
Rotulador-PJe: Ferramenta para anotação de dados em processos judiciais trabalhistas  Obs.: trabalho inscrito na política de ações afirmativas	Rodrigo Cartaxo Marques Duarte; Hildeberto Abreu Magalhaes
"Impacto da Governança de Dados na Justiça do Trabalho: Análise dos Indicadores de Desempenho e Boas Práticas de Gestão no TRT13"	Fernanda Silva de Lima; Otaviano José do Nascimento Alcântara; Larissa Leonia Bezerra de Andrade Albuquerque
Sabiá - Análise de Similaridade em Recursos Trabalhistas: Uma Abordagem de Aprendizado de Máquina e Ciência de Dados	Eduardo Camargo de Siqueira; Nilton Correia da Silva; Eduardo Ramos Correa
PIETI: Integração de Dados na Luta contra o Trabalho Infantil  Obs.: trabalho inscrito na política de ações afirmativas	Christiane Purificação de Castro; Stephanie Almeida de Jesus Dias; Renato Máximo Sátiro
Cruzamento das bases Codex e RAIS para análise do perfil dos trabalhadores que acessam a Justiça do Trabalho	Alexandre Samy de Castro; Alessandro Vinicius Amaral de Moura Beltrão; Jefferson Silva dos Anjos; Miguel Nathan Foguel; Carlos Henrique Corseuil
Similaridade por assunto nos Despachos de Admissibilidade de processos trabalhistas	Aline Dayany Lemos; Camila Ribeiro Rocha Torres; Ana Carolina Pereira Rocha

---

### **TRABALHO 1**

**TÍTULO: Rotulador-PJe: Ferramenta para anotação de dados em processos judiciais trabalhistas**

## **AUTORES**

Rodrigo Cartaxo Marques Duarte, Tribunal Regional do Trabalho da 13ª Região

Hildeberto Abreu Magalhaes, Tribunal Regional do Trabalho da 13ª Região

## **QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**

O problema principal que a solução de rotulagem adaptada aos processos judiciais pretende atacar é grande quantidade de cliques presente no processo repetitivo de anotação de dados, bem como a dificuldade de acompanhamento de todo o processo, desde a quantidade de processos rotulados, a produtividade de cada especialista e a identificação de insights parciais durante o processo.

## **MÉTODOS E TÉCNICAS UTILIZADOS**

O desenvolvimento da solução de rotulagem de processos se deu a partir da adaptação de solução open-source à realidade da Justiça do Trabalho, integrando-a ao banco de dados do PJe e estabelecendo rotulos e procedimentos que aceleram o processo como um todo. Foi concebida, inicialmente, uma infraestrutura de data lake por meio da qual os dados são extraídos do PJe, tratados e disponibilizados. Paralelamente, foi estabelecido um conjunto de rótulos relacionados à estrutura da petição inicial e da contestação, as quais conterão as informações necessárias para treinamento futuro dos modelos de IA. E como última etapa do desenvolvimento, foi desenvolvido um frontend adaptado de uma solução open-souce, idealizada para tornar o processo o mais suave e intuitivo possível.

## **CONCLUSÕES:**

Como resultados deste projeto, foi disponibilizado em ambiente de produção a primeira versão do software, que será utilizado para rotulagem de petições iniciais e contestações a partir de Agosto de 2024. Os resultados finais do programa de rotulagem, que diz respeito à criação da base de dados anotados e ao treinamento de modelos de IA supervisionados, serão etapas posteriores com o uso desta ferramenta.

---

## **TRABALHO 2**

**TÍTULO: Impacto da Governança de Dados na Justiça do Trabalho: Análise dos Indicadores de Desempenho e Boas Práticas de Gestão no TRT13**

## **AUTORES**

Fernanda Silva de Lima – Tribunal Regional do Trabalho da 13ª Região

Otaviano José do Nascimento Alcântara - Tribunal Regional do Trabalho da 13ª Região

Larissa Leonia Bezerra de Andrade Albuquerque - Tribunal Regional do Trabalho da 13ª Região

### **QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**

Este estudo destaca a importância da governança de dados na Justiça do Trabalho, demonstrando que a gestão eficiente dos dados é um fator crucial para melhorar a qualidade e a eficiência dos serviços prestados à sociedade.

Questão/Problema/Proposta Central

1. Longos prazos processuais: Desde o ajuizamento da ação até a realização da 1ª audiência e até o julgamento.
2. Inconsistências no lançamento de movimentações no sistema PJe: O que resulta em dados estatísticos imprecisos.
3. Impacto nos indicadores de desempenho do IGEST e metas do CNJ: Especialmente nos mesoindicadores do IGEST Acervo, Celeridade, Produtividade, Congestionamento e Força de Trabalho e Meta 1 do CNJ.
4. Necessidade de melhoria na gestão de unidades: Para otimizar o trabalho e cumprir as metas estabelecidas pelo CNJ.

Justificativa A governança de dados eficiente é crucial para melhorar a celeridade e a qualidade dos processos judiciais. Dados precisos e bem geridos permitem uma análise mais eficaz dos prazos processuais, identificando gargalos e implementando boas práticas para otimizar o desempenho das unidades jurisdicionais. Com a correta inserção de dados no sistema PJe é possível extrair informações estatísticas confiáveis pelo sistema e-Gestão, contribuindo para a tomada de decisões mais assertivas e para a melhoria contínua da gestão.

Além disso, uma boa governança de dados impacta diretamente na colocação das unidades no Índice Nacional de Gestão do Desempenho da Justiça do Trabalho (IGEST) e no cumprimento das metas do CNJ, resultando em maior transparência e eficiência no serviço prestado à sociedade.

### **MÉTODOS E TÉCNICAS UTILIZADOS**

1. Coleta e Análise de Dados: Utilização de ferramentas estatísticas para analisar os prazos médios processuais desde o ajuizamento da ação até o julgamento. A análise incluirá: Prazos médios da fase de conhecimento. Quantitativo de processos solucionados em relação ao total de casos novos. Quantidade de processos baixados no período e o Índice de atendimento à Demanda (IAD).
2. Ferramentas de Visualização de Dados: Criação de gráficos para apresentar os prazos médios e identificar gargalos processuais.

3. 3. Comparação com Indicadores de Desempenho: Avaliação dos dados em relação aos mesoindicadores do IGEST (Acervo, Celeridade, Produtividade, Congestionamento e Força de Trabalho) e metas do CNJ.
4. 4. Revisão de Boas Práticas: Compartilhamento de boas práticas no sistema PJe para melhorar a gestão das unidades e otimizar o trabalho.
5. 5. Reuniões: Realização de reuniões/treinamentos para magistrados e servidores sobre o lançamento correto das movimentações no sistema PJe e a importância de dados precisos para a governança de dados.

## **CONCLUSÕES:**

### Conclusões Parciais:

1. Identificação de gargalos nos prazos processuais e inconsistências nos dados lançados no sistema PJe.
2. Conscientização de magistrados e servidores sobre a importância da correta inserção de dados.
3. Implementação de algumas boas práticas de gestão, resultando em melhorias iniciais nos indicadores de desempenho.

### Conclusões Finais (ESPERADAS):

1. Melhoria significativa nos prazos médios processuais, contribuindo para o aumento da celeridade.
2. Dados estatísticos mais precisos e confiáveis, possibilitando uma gestão mais eficaz e transparente.
3. Aumento no cumprimento das metas do CNJ e melhor posicionamento das unidades no IGEST.
4. Cultura de governança de dados fortalecida, com magistrados e servidores engajados na busca contínua por melhorias.

---

## **TRABALHO 3**

**TÍTULO: Sabiá - Análise de Similaridade em Recursos Trabalhistas: Uma Abordagem de Aprendizado de Máquina e Ciência de Dados**

### **AUTORES**

Eduardo Camargo de Siqueira; Instituto Federal de Brasília; DF

Nilton Correia da Silva; Universidade de Brasília; DF

Eduardo Ramos Correa; Departamento de Ciência de Dados TST; DF

### **QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**

Neste trabalho, foi utilizado o modelo Sentence-BERT (REIMERS e GUREVYCH, 2019), pré-treinado com dados jurídicos, para encontrar sentenças semelhantes

em peças de Recursos de Revista (RR). Sentence-BERT é uma técnica de aprendizado de máquina baseada em transformadores que permite capturar a semântica das sentenças e 6 medir a similaridade entre elas de forma eficaz. Esse modelo foi aplicado para analisar a complexidade dos casos trabalhistas e identificar padrões de similaridade, mesmo quando os recursos apresentam reivindicações diversificadas. Além disso, foram integradas heurísticas específicas do domínio jurídico para aprimorar a precisão das correspondências e facilitar o processo de identificação de precedentes legais.

## **MÉTODOS E TÉCNICAS UTILIZADOS**

O TST-Sabiá engloba a pesquisa de métodos de aprendizado de máquina (AM) para enfrentar o desafio da similaridade entre recursos do Tribunal Superior do Trabalho do Brasil. A complexidade dos casos trabalhistas adiciona uma camada de dificuldade, pois casos podem ser semelhantes em certos aspectos, enquanto distintos em outros. O contexto brasileiro, com uma carga processual significativa, apresenta oportunidades para a aplicação de Inteligência Artificial (IA) no sistema judicial. Neste sentido, DA SILVA, et. al. (2018) apresentam suas descobertas sobre a identificação de peças processuais para o Supremo Tribunal Federal do Brasil usando uma simples rede neural convolucional. DEEKS (2019) enfatiza a importância de que os algoritmos sejam transparentes para as partes envolvidas no processo legal. Por isso, este artigo apresenta uma abordagem utilizando AM e técnicas de ciência de dados para identificar recursos semelhantes, seja para fins de votação ou na criação de precedentes.

## **CONCLUSÕES:**

A aplicação do modelo Sentence-BERT, pré-treinado com dados jurídicos, demonstrou eficácia na identificação de sentenças semelhantes em peças de Recursos de Revista (RR). A capacidade do modelo de capturar nuances semânticas entre textos jurídicos complexos mostrou-se promissora para a padronização e agilização dos processos no Tribunal Superior do Trabalho. A integração de heurísticas específicas do domínio contribuiu significativamente para a precisão das correspondências. Esses resultados indicam um avanço relevante na análise de similaridade legal, facilitando a criação de precedentes e potencialmente melhorando a eficiência e a justiça no julgamento dos recursos trabalhistas.

---

## **TRABALHO 4**

### **TÍTULO: PIETI: Integração de Dados na Luta contra o Trabalho Infantil**

#### **AUTORES**

Christiane Purificação de Castro - Tribunal Regional do Trabalho da 6ª Região;

Stephanie Almeida de Jesus Dias - Tribunal Regional do Trabalho da 6ª Região;

### **QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**

O trabalho infantil é um problema global que afeta milhões de crianças, comprometendo seu desenvolvimento físico, moral e educacional. Apesar dos avanços em políticas públicas, a persistência desse problema no Brasil demanda abordagens inovadoras que possam identificar e combater suas causas profundas de maneira mais eficaz. Este estudo investiga os mecanismos institucionais de combate ao trabalho infantil no âmbito do judiciário trabalhista, com foco na integração e análise de dados. Propomos a criação da Plataforma Integrada de Erradicação do Trabalho Infantil (PIETI), que utiliza ciência de dados para consolidar informações provenientes de diversas fontes, como registros judiciais e dados socioeconômicos. A PIETI visa fortalecer as intervenções existentes, melhorar a eficácia das políticas públicas e promover uma resposta mais coordenada e baseada em evidências. A plataforma permitirá a identificação de padrões e tendências, facilitando a formulação de estratégias preventivas e corretivas visando a erradicação do trabalho infantil.

### **MÉTODOS E TÉCNICAS UTILIZADOS**

O estudo adota uma abordagem multimetodológica composta por três fases: (1) exploratória-descritiva, compreendendo o cenário atual do trabalho infantil e identificando fontes de dados relevantes, incluindo dados de processos judiciais e indicadores socioeconômicos; (2) descritiva, coletando, estruturando e analisando dados quantitativos, utilizando, entre outras técnicas, mapas de calor e análise de clusters para visualizar áreas de maior incidência e identificar padrões; (3) implementação e monitoramento, desenvolvendo a Plataforma Integrada de Erradicação do Trabalho Infantil (PIETI) com tecnologias de Big Data e machine learning para monitorar, ajustar e avaliar continuamente as políticas públicas.

### **CONCLUSÕES:**

Os resultados iniciais indicam que a integração de dados judiciais e socioeconômicos pode melhorar a identificação dos casos de trabalho infantil, contribuindo para a construção de ações estratégicas preventivas visando sua erradicação. A utilização da PIETI permitiu identificar padrões, possibilitando intervenções mais precisas e demonstrando a importância de registros adequados e completos para a identificação da incidência e o tratamento adequado dos casos de trabalho infantil. Como principais achados, identificou-se que o ramo de atividade e informações constantes nas características processuais específicas influenciam a incidência de trabalho infantil. Os resultados parciais demonstram a necessidade de avançar nas investigações e refinar a definição do que constitui trabalho infantil utilizada neste estudo, assegurando que todas as formas sejam corretamente identificadas e abordadas.

## **TRABALHO 5**

### **TÍTULO: Cruzamento das bases Codex e RAIS para análise do perfil dos trabalhadores que acessam a Justiça do Trabalho**

#### **AUTORES**

Alexandre Samy de Castro - Técnico de Planejamento e Pesquisa do IPEA

Alessandro Vinicius Amaral de Moura Beltrão - Tribunal Superior do Trabalho Brasília/DF

Jefferson Silva dos Anjos - Bolsista do PNPd/IPEA

Miguel Nathan Foguel - Técnico de Planejamento e Pesquisa do IPEA

Carlos Henrique Corseuil - Técnico de Planejamento e Pesquisa do IPEA

#### **QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**

Encontra-se atualmente em curso a pesquisa “Perfil e trajetória dos trabalhadores e empresas litigantes na Justiça do Trabalho”, em uma parceria entre o Instituto de Pesquisa Econômica Aplicada (IPEA) e a Justiça do Trabalho (representada pela Secretaria de Pesquisa Judiciária e Ciência de Dados do Tribunal Superior do Trabalho). O objetivo da pesquisa é a produção de análise estatística, de caráter descritivo, da relação entre os processos trabalhistas existentes no Brasil entre 2016 e 2023 e o perfil e trajetórias dos dois agentes que compõem esses litígios: o trabalhador e a firma. Este trabalho se propõe a descrever as bases de dados utilizadas nesta análise, e como foram realizadas as etapas de extração e cruzamento dos dados.

#### **MÉTODOS E TÉCNICAS UTILIZADOS**

Na pesquisa são utilizados os dados de duas bases: a plataforma Codex, instituída pela Resolução nº 446 de 14/03/2022 do Conselho Nacional de Justiça (CNJ) como ferramenta oficial de extração de dados estruturados e não estruturados dos processos judiciais eletrônicos em tramitação no Poder Judiciário Nacional; e a Relação Anual de Informações Sociais (RAIS), regida atualmente pelo Decreto nº 10.854, de 10/11/2021, que contém uma série de informações laborais de trabalhadores e firmas formais no Brasil. Os dados da plataforma Codex foram extraídos a partir da API disponibilizada pelo CNJ para acesso aos dados dos tribunais trabalhistas. Os dados extraídos, no formato JSON, foram enviados ao IPEA, que realizou o tratamento dos dados, transformando-os em um banco relacional em MySQL. Foi feita também uma meta-análise dos dados recebidos, para validação da consistência das informações. O cruzamento com a RAIS é feito a partir do CPF do reclamante. A RAIS é um banco com cerca de 700 milhões de registros entre 2011 e 2021. Para cada CPF encontrado no cruzamento, guarda-se todas as informações de vínculos formais, independentemente de se a empresa que consta do vínculo ser a mesma empresa que figura no polo passivo da reclamação trabalhista.

## **CONCLUSÕES:**

Cruzamentos preliminares, indicam que 94% dos CPFs de reclamantes (entre 2016 e 2023) aparecem pelo menos uma vez na RAIS (entre 2011 e 2021). O cruzamento é complexo devido ao tamanho das bases de dados. Por um lado, temos cerca de seis milhões e meio de processos judiciais. De outro lado, temos cerca de 770 milhões de registros de vínculos formais de emprego, na base RAIS. Além disso, para cada CPF, podemos ter múltiplos processos e múltiplos vínculos de emprego ao longo do tempo. Para cada processo judicial, por sua vez, podemos ter múltiplos CPFs e múltiplos CNPJ/raiz distintos. Diante desta multiplicação de registros, a análise descritiva dos dados requer cuidados especiais com relação às unidades de observação (que podem ser o processo, o CPF, o CNPJ, etc.) e com relação a unicidade dos registros, no sentido de se evitar redundâncias no dados, que possam resultar em distorções na análise. O cruzamento é feito em duas etapas: 1- cruzamento utilizando apenas o CPF como chave; 2- cruzamento para identificar em quais processos o CPF se encontra pareado com o CNPJ, tanto no processo quanto no vínculo de emprego formal. Resultados preliminares sugerem que os reclamantes não pareados, isto é aqueles que não tiveram nenhum vínculo de emprego formal com a reclamada, tem características bastante distintas daqueles reclamantes pareados (aqueles que estão processando empresas com as quais tiveram vínculo formal de emprego). Estas diferenças compreendem não somente as características do processo, mas também o perfil dos trabalhadores, do ponto de vista de gênero, escolaridade, rotatividade, etc. Um segundo exercício consiste em comparar os reclamantes formais com os não-reclamantes formais. Esta análise parte do sorteio de uma amostra de trabalhadores formais na RAIS, que jamais ajuizaram uma reclamação trabalhista. Comparando às características de litigantes e não-litigantes, observamos também diferenças significativas no perfil destes dois grupos de trabalhadores formais. Os resultados preliminares servirão a um debate mais amplo sobre o acesso à justiça. Dentre as questões relevantes, devemos destacar a questão das relações de trabalho informal e de que maneira esta dinâmica econômica se reflete sobre as demandas apresentadas à Justiça do Trabalho.

---

## **TRABALHO 6**

**TÍTULO: Similaridade por assunto nos Despachos de Admissibilidade de processos trabalhistas**

### **AUTORAS**

Aline Dayany Lemos - Universidade de Brasília;

Camila Ribeiro Rocha Torres - Tribunal Superior do Trabalho

Ana Carolina Pereira Rocha - Tribunal Superior do Trabalho

**QUESTÃO/ PROBLEMA/PROPOSTA CENTRAL:**



O Tribunal Superior do Trabalho (TST), diante de um acervo crescente e da urgência por acelerar os tempos de julgamentos de processo, fez uma parceria com a Universidade de Brasília (UnB) e está concluindo o desenvolvimento do projeto Sabiá, que tem dentre seus objetivos encontrar processos similares no TST a partir da análise das peças processuais, identificando as similaridades dos processos para cada assunto debatido. Neste trabalho, apresenta-se a aplicação de modelos de processamento de linguagem natural e um algoritmo de agrupamento para encontrar os grupos de processos semelhantes dentro de um mesmo assunto processual por meio do Despacho de Admissibilidade.

## **MÉTODOS E TÉCNICAS UTILIZADOS**

A partir de uma análise quantitativa e qualitativa, buscou-se entender quais eram as classes de processos mais numerosos no acervo do TST e quais seriam os tipos de peças mais relevantes para ajudar a encontrar a similaridade entre os processos, considerando a pluralidade de assuntos de um processo. Dentre as peças escolhidas tem-se o Despacho de Admissibilidade, onde consta a análise tema a tema feita pelos Tribunais Regionais do Trabalho a respeito do cabimento da análise do Recurso de Revista pelo TST. Esta peça é frequentemente redigida por meio da ferramenta e-Rec, que guarda informações estruturadas de cada tema do processo. Quando esta ferramenta não é utilizada, tem-se apenas o texto corrido, o que dificulta a separação das informações de cada tema. Utilizando-se os modelos Bert1 e K-Means2, o texto é processado e então agrupado dentro de cada assunto, permitindo encontrar nuances de discussões dentro de um mesmo assunto e a indicação dos processos mais similares na temática.

## **CONCLUSÕES:**

Para os documentos não originários do eRec, ou seja, os documentos que são dispostos como texto corrido, apesar de ter sido possível segmentar e identificar um assunto e o texto correspondente, não foi possível relacionar os assuntos aos assuntos da Tabela Unificada de Assuntos do CNJ dentro desta pesquisa, uma vez que não havia uniformidade na redação dos mesmos, assim, estes documentos não passaram a etapa seguinte de agrupamento, e deixaram de ser escopo do Projeto Sabiá. Para os demais documentos, provenientes do eRec, os grupos dentro de cada assunto foram formados, apresentando uma mediana de 95% de similaridade dentro de um mesmo grupo. A inferência da similaridade e os grupos formados foram avaliados de diferentes formas pela equipe negocial. Resultados preliminares apresentam parecer favorável e indicam que as sugestões fornecidas para processos similares em um mesmo tema na peça de Despacho de Admissibilidade estão apropriadas.