

DADOS DE FONTES EXTERNAS PARA O APRIMORAMENTO DA FISCALIZAÇÃO TRABALHISTA NO BRASIL

Enadio da Silva Barbosa¹
Jefferson de Moraes Toledo²
Magno Cavalcante da Silva³
Robson Dias Alves Timoteo⁴

1. Introdução. 2. Dados de fontes externas à Inspeção do Trabalho. 2.1. Dados do IBGE. 2.2. Dados de outras fontes. 3. O sistema. 4. Usos do sistema *Pool of Features*. 5. Considerações finais.

RESUMO

A transformação digital ocorrida no mundo nas últimas décadas continua repercutindo em todas as esferas da sociedade, inclusive nos diversos ramos e níveis da Administração Pública. As consequências naturais desse processo foram, por um lado, uma crescente disponibilidade de dados e, por outro, um aumento significativo da necessidade de utilizá-los em prol da melhoria na prestação de serviços, quer seja em âmbito público, quer seja na esfera privada. No que tange à Inspeção do Trabalho brasileira, essas mudanças não passam despercebidas e várias iniciativas começam a tomar forma. Porém, nesse caso, uma grande quantidade de dados precisa estar meticulosamente organizada para que possa servir de suporte e exercer a utilidade que dela se espera. Nesse diapasão, o presente estudo propõe o Projeto *Pool of Features* (“Piscina de Características”, em tradução literal) que representa o esforço de trazer, para o ambiente informático da Inspeção do Trabalho, de forma estruturada, dados de fontes externas que possam contribuir na melhoria da qualidade da prestação do serviço público realizado pela Fiscalização Trabalhista, impulsionando o planejamento de ações baseado em inteligência fiscal.

Palavras-chave: Governo digital. Dados públicos. Planejamento fiscal.

¹ Auditor-Fiscal do Trabalho. Graduado em Direito e Ciências Econômicas. Pós-graduado em engenharia de software e MBA internacional pelo instituto de empresa - Espanha.

² Auditor-Fiscal do Trabalho. Doutor em física, com ênfase na área de gravitação e cosmologia.

³ Auditor-Fiscal do Trabalho.

⁴ Auditor-Fiscal do Trabalho. Mestre e doutor em Ciência da Computação pela Universidade Federal de Pernambuco.

1 INTRODUÇÃO

Embora a utilização de tecnologia por governos não seja recente, esse fenômeno vem adquirindo novas características e mudando por completo a relação entre Estado e sociedade (FILGUEIRAS; FLÁVIO; PALOTTI, 2019; VARELLA; OLIVEIRA; MOESCH, 2017). Os gestores públicos têm precisado lidar com questões como a necessidade de fornecimento de serviços digitais, formulação de políticas públicas com base em dados, operação de programas sociais em ambientes virtuais etc. (JANOWSKI, 2015). Tal movimento tem levado a uma verdadeira transformação da atividade governamental em si. Muitos autores têm se debruçado sobre o tema da evolução do governo digital e há quem estabeleça uma sistematização do processo (FLAK; MOE; SÆBØ, 2003; JANOWSKI, 2015; FALK; RÖMMELE; SILVERMAN, 2017). É possível, por exemplo, sistematizar a evolução do fenômeno nas seguintes etapas (JANOWSKI, 2015):

- **DIGITALIZAÇÃO:** a tecnologia começa a ser usada no governo;
- **TRANSFORMAÇÃO:** a tecnologia começa a impactar a própria organização do governo;
- **ENGAJAMENTO:** a tecnologia começa a influenciar agentes externos e há o fornecimento de serviços digitais à sociedade;
- **CONTEXTUALIZAÇÃO:** a digitalização impacta a criação de políticas públicas e impacta de forma efetiva a sociedade.

É fato que essas classificações são didáticas e não abordam, de forma exaustiva, o complexo tema da digitalização governamental, assim como não há uma sequência linear entre as etapas apresentadas, podendo estas coexistirem e até mesmo fundir-se umas nas outras. É fundamental, entretanto, destacar que o presente trabalho tem o objetivo de ser uma importante ferramenta para a inserção da Inspeção do Trabalho na fase de contextualização e, como será discutido mais adiante, permitir o desenvolvimento e adoção de políticas públicas de fiscalização com base em variáveis capturadas digitalmente da sociedade.

Diversos órgãos governamentais do Brasil dos níveis municipal, estadual e federal têm experimentado a transformação digital de seus serviços (FILGUEIRAS; FLÁVIO; PALOTTI, 2019). Os esforços para a informatização de serviços públicos

federais remontam à década de 1960 e, mais recentemente, tem-se experimentado etapas de verdadeira revolução digital.

No que concerne à Inspeção do Trabalho, é possível verificar que há, tanto no Brasil quanto fora dele, uma necessidade de inserção no mundo digital visando à melhoria dos impactos sociais de suas ações (SUZUKI, 2016; MUSTCHIN; LUCIO, 2020; DAHL; STARREN, 2019; TOLEDO; TIMOTEO; BARBOSA, 2020). No Brasil, por exemplo, já existem exemplos de transformação digital (como o sistema de lavratura de documentos fiscais, Sistema Auditor, o sistema de fiscalização de jornada laboral, Khronos, e o sistema de gerenciamento de fiscalizações, SFITWEB) e de engajamento digital (como o sistema de processo eletrônico de autos de Infração e de notificações de débito de FGTS, que permite interação com os administrados).

Neste estudo, propõe-se um sistema que permite a organização e estruturação de dados advindos de fontes externas à Inspeção do Trabalho, de forma que sejam possíveis a rápida geração de *dashboards* (tela de apresentação de informações) e a disponibilização de *features* (características) que possam alimentar modelos preditivos e regressivos de *machine learning*.

A Lei nº 10.593, de 6 de dezembro de 2002 (BRASIL, 2002a), e o Regulamento da Inspeção do Trabalho (BRASIL, 2002b) conferiram aos Auditores-Fiscais do Trabalho suas respectivas competências e atribuições. Dentre elas, é possível listar: assegurar o cumprimento da legislação trabalhista, incluindo as condições de segurança e saúde no trabalho; a verificação da assinatura das Carteiras de Trabalho e Previdência Social (CTPS) dos trabalhadores, visando à redução da informalidade; a verificação dos recolhimentos ao Fundo de Garantia do Tempo de Serviço (FGTS); o combate ao trabalho infantil ilícito; a inserção de pessoas com deficiência no mercado de trabalho; e o combate ao trabalho análogo ao de escravo.

Verifica-se, portanto, que a Fiscalização Trabalhista no Brasil possui diversas competências distintas, o que torna a etapa de planejamento das ações crucial e desafiadora para a consecução de uma melhoria contínua na prestação do serviço público. Para algumas áreas, como a fiscalização dos recolhimentos do FGTS, os dados disponíveis permitem a realização de cruzamentos e determinação de indícios de irregularidade, facilitando o processo de escolha de empresas para auditoria. Em outras áreas, entretanto, há grande necessidade de avançar no sentido do

aperfeiçoamento e intensificação do uso de dados no processo de planejamento das ações. Para a fiscalização de segurança e saúde no trabalho, é possível, por exemplo, utilizar dados de inspeções pretéritas para realizar predições da ocorrência de acidentes, de modo que se auxilie o caráter preventivo da Inspeção do Trabalho, redundando na diminuição da perda de vidas e na redução do sofrimento humano causado pela ocorrência de acidentes laborais, bem como na economia de recursos públicos e decréscimo dos impactos negativos nas atividades empresariais (TOLEDO; TIMOTEO; BARBOSA, 2020).

É fundamental destacar que o presente trabalho representa uma iniciativa inovadora da Inspeção do Trabalho, propondo uma sistemática abrangente de agregação de dados fundamentais aos trabalhos desempenhados pela categoria e não se mostra como um fim em si mesmo. Pelo contrário, a proposta dos autores é de que os dados trazidos pelo sistema sugerido possam auxiliar no cumprimento dos seguintes objetivos, entre outros:

- **PLANEJAMENTO DAS AÇÕES DE FISCALIZAÇÃO:** Utilizar intensivamente informações na etapa de planejamento, de forma a melhorar o resultado das ações e aprimorar a atuação dos Auditores-Fiscais do Trabalho.

- **CONSTRUÇÃO DE MODELOS PREDITIVOS DE CLASSIFICAÇÃO E REGRESSÃO:** Alimentar com *features* (características) modelos de *machine learning*, de maneira que os dados permitam intensificar o trabalho preventivo da Auditoria-Fiscal do Trabalho, mitigando os efeitos deletérios de infrações trabalhistas para os empregados e para a sociedade.

- **AUMENTO DA DISSEMINAÇÃO DE INFORMAÇÕES PARA O TRABALHO DIÁRIO DOS AUDITORES:** Alimentar *dashboards* (painéis estatísticos) que podem ser disponibilizados para chefias e Auditores, com o intuito de auxiliar o trabalho nas ações fiscais.

- **AVALIAÇÃO DO IMPACTO DA FISCALIZAÇÃO TRABALHISTA NA SOCIEDADE:** Verificar a efetividade das ações da Auditoria-Fiscal do Trabalho por meio de variáveis como taxa de informalidade, proporção de crianças/adolescentes em condições de trabalho irregular etc.

Este artigo está organizado como segue. Na seção 2, serão apresentados os dados externos que foram incluídos nos estudos realizados, inclusive citando as

fontes e formas de acesso. Serão apresentados, na seção 3, o desenho do sistema desenvolvido para automatizar a extração e o tratamento dos dados de fontes externas, bem como a tecnologia utilizada em sua construção. Na seção 4, serão abordadas as aplicações correntes e futuras dos dados trazidos pela Fiscalização Trabalhista, com destaque à construção de modelos de inteligência artificial para auxílio ao planejamento. Por fim, na seção 5, serão apresentadas as considerações finais.

2 DADOS DE FONTES EXTERNAS À INSPEÇÃO DO TRABALHO

Como se tem discutido, a Inspeção do Trabalho no Brasil necessita realizar planejamento de suas ações com o foco na ampliação da efetividade de seu alcance, promovendo verdadeira transformação social no País. Para tanto, faz-se indispensável não apenas utilizar dados de posse do Ministério do Trabalho e Previdência Social (MTP) ou produzidos pela Auditoria-Fiscal do Trabalho, mas buscar informações externas, principalmente originadas de pesquisas estatísticas.

Nesta seção, discorrer-se-á sobre os dados públicos externos à Fiscalização do Trabalho que foram importados no escopo deste trabalho, apresentando-se suas fontes, metodologia e formas de acesso.

Vale destacar que os dados aqui utilizados neste trabalho são selecionados de tal forma que seja possível calcular variáveis estatísticas relacionadas ao mundo do trabalho ou, alternativamente, que tenham alguma possível relevância para a construção de modelos de inteligência artificial. Importa, ainda, destacar que, como a Fiscalização do Trabalho tem a responsabilidade de verificar o cumprimento das normas trabalhistas em todo o território nacional, é fundamental que as variáveis externas importadas permitam a construção de soluções com o menor nível de agregação territorial possível, e que sejam recentes e periodicamente atualizadas.

2.1 DADOS DO IBGE

O Instituto Brasileiro de Geografia e Estatística - IBGE é uma entidade pública vinculada ao Ministério da Economia e se "constitui no principal provedor de dados e

informações do País, que atendem às necessidades dos mais diversos segmentos da sociedade civil, bem como dos órgãos das esferas governamentais federal, estadual e municipal" (IBGE, 2021a).

O instituto realiza a produção de divulgação de pesquisas estatísticas, assim como a consolidação e disponibilização dos resultados. As informações divulgadas pelo IBGE abrangem diversos temas como população, trabalho, educação, saúde e estatísticas econômicas. Muitos dados são importantes à Inspeção do Trabalho, permitindo um acompanhamento de variáveis relacionadas ao mercado de trabalho, servindo de base para a construção de indicadores de planejamento ou auxiliando na construção de modelos de *machine learning* (aprendizado de máquina em tradução livre), área de estudos da inteligência artificial.

O IBGE faz a divulgação de dados e resultados por meio de diversos canais. Um deles é o Sistema IBGE de Recuperação Automática - SIDRA (IBGE, 2021b), através do qual as tabelas derivadas de pesquisas e já sumarizadas pelo IBGE estão disponíveis e é possível consultar a documentação e fazer *download* dos dados. O mesmo sistema disponibiliza uma API (*Application Programming Interface* - Interface de programação de aplicações), que permite a utilização dos dados da pesquisa por outras aplicações. No presente trabalho, essa API é utilizada para automatizar as extrações dos dados disponíveis no SIDRA. Os dados estão disponíveis por intermédio de tabelas listadas no próprio sítio do sistema na Internet (IBGE, 2021b) e, na sequência, serão enumerados os dados importados com a ajuda desta API, tais como informações gerais dos municípios do Brasil e dados de pesquisas econômicas, populacionais e da força de trabalho.

Importante destacar que a área territorial dos municípios é uma grandeza fundamental para o cálculo de variáveis derivadas, como a taxa de povoamento de uma região. Além disso, algumas tentativas já realizadas pela Subsecretaria de Inspeção do Trabalho (SIT) para distribuir a força fiscal pelo País levaram em consideração essa variável. Por esse motivo, é de fundamental relevância buscar os dados de área territorial no sistema.

No que se refere aos dados populacionais, a estimativa da população dos municípios do País é realizada pelo IBGE (OLIVEIRA; SIMÕES, 2005; BORGES; ERVATTI; SILVA, 2011). Essa é uma grandeza de fundamental importância em vários

trabalhos atuais e futuros da Auditoria-Fiscal do Trabalho, já que há variáveis relacionadas ao mundo do trabalho que com ela guardam importante relação, como a população economicamente ativa e a população ocupada. Outro uso importante dessa variável é no cálculo de variáveis derivadas essenciais para a comparação de unidades territoriais que se diferenciam por suas dimensões e características, como o Produto Interno Bruto (PIB) *per capita*.

Por seu turno, o PIB de uma unidade territorial mede a soma de todos os bens e produtos finais produzidos na região (IBGE, 2021). Essa é uma grandeza de fundamental importância na mensuração da atividade econômica de determinado território e pode ser trazida da tabela de número 5938 do SIDRA.

O IBGE também mantém o Cadastro Central de Empresas (Cempre), compilado de informações obtidas por meio de dados da Secretaria de Trabalho do MTP, como aqueles oriundos do Relatório Anual de Informações Sociais (RAIS) e do Cadastro Geral de Empregados e Desempregados (CAGED), e de pesquisas anuais realizadas pelo IBGE, como a Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) (IBGE, 2018). Esse cadastro traz dados fundamentais à Inspeção do Trabalho, a exemplo do número de empresas em cada nível de agregação territorial, da quantidade de pessoas ocupadas e assalariadas.

Na tabela 1, resumam-se as grandezas importadas a partir da API do SIDRA, informando-se o número da tabela de origem dos dados e, quando importante, o número da variável correspondente à grandeza entre parênteses. Na última coluna da tabela, informa-se o mínimo nível de agregação da variável importada.

Tabela 1 – Dados do IBGE importados para uso pela Inspeção do Trabalho

Grandeza	Fonte	Nível de agregação
Área territorial	Tabela 6579 (IBGE)	Município
População estimada	Tabela 1301 (IBGE)	Município
PIB	Tabela 5938 (IBGE)	Município
Número de Empresas	Tabela 6449 (IBGE)	Município – Seção do CNAE
Pessoal Ocupado	Tabela 6449 (IBGE)	Município – Seção do CNAE
Pessoal Assalariado	Tabela 6449 (IBGE)	Município – Seção do CNAE

Fonte: Elaborada pelos autores com base em dados obtidos da API/SIDRA do IBGE

Embora, como já se afirmou, o SIDRA sintetize dados da PNADc, há informações essenciais derivadas da pesquisa que precisam ser calculadas com base em seus microdados (dados com informações dos questionários da pesquisa). É o caso, por exemplo, dos dados de informalidade no mercado de trabalho no País e de trabalho infantil, fundamentais para o exercício das competências da Fiscalização do Trabalho.

A PNADc é uma pesquisa amostral desenvolvida pelo IBGE com o objetivo de acompanhar, com periodicidade trimestral, dados da força de trabalho no País (IBGE, 2021c). Diferentemente dos dados tratados até agora, a PNADc não produz resultados para todos os municípios do Brasil, tendo abrangência territorial definida pelo IBGE, conforme se lê a seguir (IBGE, 2021c):

Sua amostra foi planejada de modo a produzir resultados para Brasil, Grandes Regiões, Unidades da Federação, Regiões Metropolitanas que contêm Municípios das Capitais, Região Integrada de Desenvolvimento - RIDE Grande Teresina, e Municípios das Capitais. Desde sua implantação, a pesquisa, gradualmente, vem ampliando os indicadores investigados e divulgados.

Nesse sentido, é fundamental agregar a este trabalho os microdados da PNADc, que precisam ser baixados em formato de arquivo texto por meio de Protocolo FTP (*File Transfer Protocol*, Protocolo de Transferência de Arquivos em tradução literal), disponibilizado pelo IBGE e com o qual também é possível realizar o *download* dos dicionários das tabelas. Com as informações da PNADc é possível, como será discutido na Seção 4, criar panoramas da informalidade no mercado de trabalho no Brasil para os níveis territoriais da pesquisa, além de elaborar modelos de inteligência artificial que realizam a predição de taxas de informalidade para municípios não abrangidos pela pesquisa.

2.2 DADOS DE OUTRAS FONTES

Além das informações trazidas a partir de pesquisas realizadas pelo IBGE, outras instituições nacionais e internacionais produzem dados que podem ser

fundamentais para os trabalhos desenvolvidos pela SIT. Listam-se, a seguir, os dados importados, com suas fontes e breve descrição de suas utilidades.

O Instituto de Pesquisa Econômica Aplicada (Ipea) é uma fundação pública ligada ao Ministério da Economia que tem como objetivo fornecer “suporte técnico e institucional às ações governamentais para a formulação e reformulação de políticas públicas e programas de desenvolvimento brasileiros” (Ipea, 2021). A partir das pesquisas do Ipea, buscou-se chegar a informações da economia dos municípios, como listado na Tabela 2, que podem constituir-se de informações importantes com correlação com a ocorrência de trabalho informal, trabalho infantil, acidentalidade laboral ou outras irregularidades trabalhistas.

Numa tentativa de alterar a visão unilateral de desenvolvimento humano relacionado apenas ao crescimento econômico de uma região, a Organização das Nações Unidas (ONU) criou o Índice de Desenvolvimento Humano (IDH) (HAQ, 1995; ONU, 2021), que avalia o desenvolvimento de um país segundo os aspectos sociais e econômicos, como qualidade de vida, renda e escolarização de sua população. Desta forma, o IDH pode guardar correlação com variáveis relacionadas ao mercado de trabalho do país.

Por fim, o nível educacional de determinada região também pode influenciar no perfil dos trabalhadores e na empregabilidade local. Sabe-se, por exemplo, que o nível de educação do trabalhador está relacionado à sua possibilidade de sofrer acidente laboral (ROCHA et al., 2019). No País, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) faz o cálculo do Índice de Desenvolvimento da Educação Básica – Ideb, que mede a qualidade educacional dos municípios, combinando-se informações de rendimento escolar (taxa de aprovação dos alunos) e desempenho em exames padronizados (Prova Brasil ou Saeb - Sistema de Avaliação da Educação Básica).

Na Tabela 2, resumem-se, esquematicamente, os dados importados a partir de fontes externas à Inspeção do Trabalho, excluindo-se o IBGE.

Tabela 2 – Dados obtidos a partir do Ipea, ONU, INEP

Grandeza	Fonte	Nível de agregação
Exportações	Ipea	Município

Importações	Ipea	Município
Receitas	Ipea	Município
Transferências correntes	Ipea	Município
Transferências de capital	Ipea	Município
IDH	ONU	Município
Ideb	INEP	Município
Prova Brasil/Saeb	INEP	Município

Fonte: Elaborada pelos autores.

3 O SISTEMA

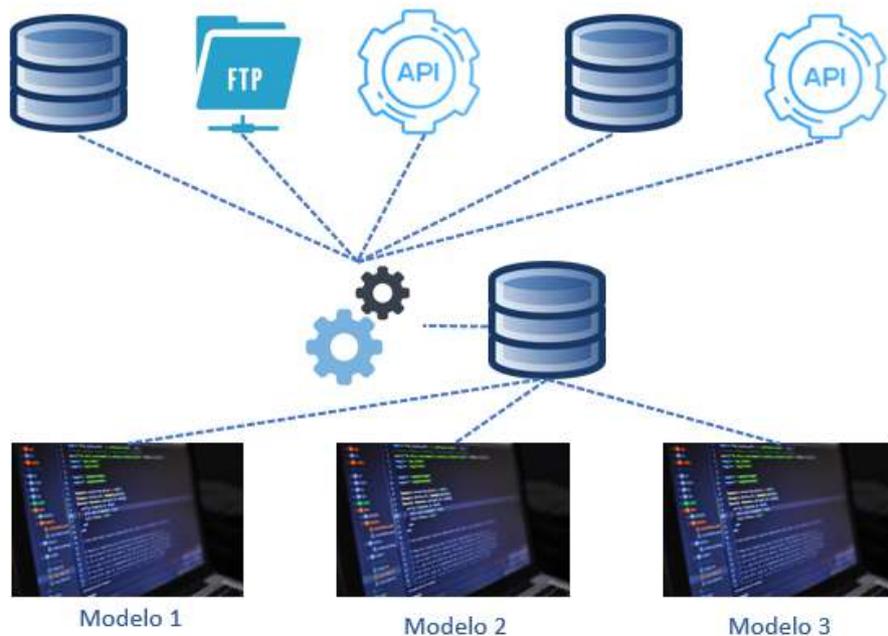
Como discorrido na seção anterior, os dados de fontes externas necessários à Inspeção do Trabalho para diversas de suas atividades precisam ser obtidos a partir de fontes distintas (API, Protocolos FTP, *download* em páginas da *Web*) disponibilizadas por diversos entes (IBGE, Ipea, ONU etc.). Além disso, há diversas informações importantes a serem acrescentadas àquelas que estão armazenadas nos bancos de dados internos do Ministério do Trabalho e Previdência, como as informações das fiscalizações trabalhistas e do eSocial, da RAIS e do CAGED.

Sendo assim, para cada aplicação ou análise que seja desenvolvida, deve ser criado todo um sistema de extração, transformação e carga dos dados (na sigla em inglês, ETL - *extract, transform and load*).

A solução proposta pela Coordenação Geral de Integração Fiscal (CGIF) da Subsecretaria de Inspeção (SIT) é a construção de sistema único que, conectando-se às diversas fontes de dados, tanto internas quanto externas à fiscalização trabalhista, possa gerar, de maneira unificada, somente uma base de dados para as diversas finalidades de planejamento e acompanhamento das ações da Inspeção do Trabalho, além de alimentar painéis estatísticos e modelos de inteligência artificial. A solução proposta é denominada de *Pool of Features*, “piscina de características” em tradução livre, já que é uma iniciativa menor, em termos de armazenamento e processamento de dados, que um *data lake*, “lago de dados” em tradução livre, nomenclatura comum aos grandes repositórios de informações. A figura 1 é uma representação esquemática do sistema proposto, mostrando a extração e o processamento unificado

de dados de diversas fontes (bancos de dados, APIs etc.) e a criação de bancos que passam a alimentar diversas iniciativas da Fiscalização do Trabalho (modelos de inteligência artificial, criação de painéis estatísticos etc.).

Figura 1 – Representação esquemática do sistema *Pool of Features*



Fonte: Elaborada pelos autores.

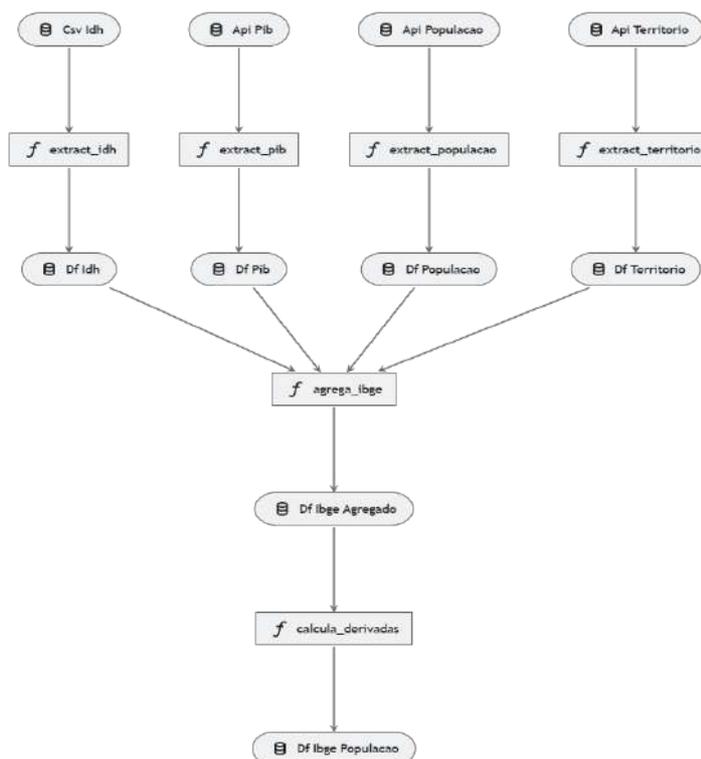
É importante destacar que a iniciativa aqui descrita abrange o sistema construído para buscar automaticamente os dados e realizar todas as transformações necessárias, mas também abarca o banco de dados gerado e que pode ser utilizado em todas as iniciativas da Fiscalização Trabalhista. A figura 1 resume o sistema, representando todas as fontes que o alimentam.

Na construção do sistema foi utilizada a linguagem de programação Python (ROSSUM et al., 1991) e a biblioteca Kedro (QUANTUMBLACK, 2021) para organizar toda a sistemática de ETL. Para manipulação dos dados são usadas as bibliotecas Pandas (MCKINNEY et al., 2011) e Spark (ZAHARIA et al., 2016). Esta última permite o trabalho com grandes volumes de dados, possibilitando o processamento paralelizado e *clusterizado* (agrupado). Além disso, o sistema é desenvolvido a partir dos conceitos de *pipelines* e nós. Um *Pipeline* é a série de transformações realizadas nos dados, desde sua extração até a geração das tabelas que serão usadas nos

diversos trabalhos desenvolvidos. Cada transformação realizada num dado é denominada de nó.

Na Figura 2, representa-se, a título de exemplo, um dos *pipelines* que compõem o sistema *Pool of Features*. Note-se que as primeiras etapas correspondem à extração dos dados de suas respectivas fontes (leitura de dados de IDH a partir de arquivo no formato csv (valores separados por vírgula), obtenção de informações de PIB, população e área territorial via APIs). Em seguida, são realizados nos nós (representados por quadriláteros) todos os cruzamentos de tabelas, agregações e demais transformações até que se tenha a geração da tabela que será usada em trabalhos futuros ou que alimentará outros *pipelines* (no exemplo da figura 2, vê-se a tabela denominada Df_ibge_População).

Figura 2 – Exemplo de *pipeline* do sistema *Pool of Features*

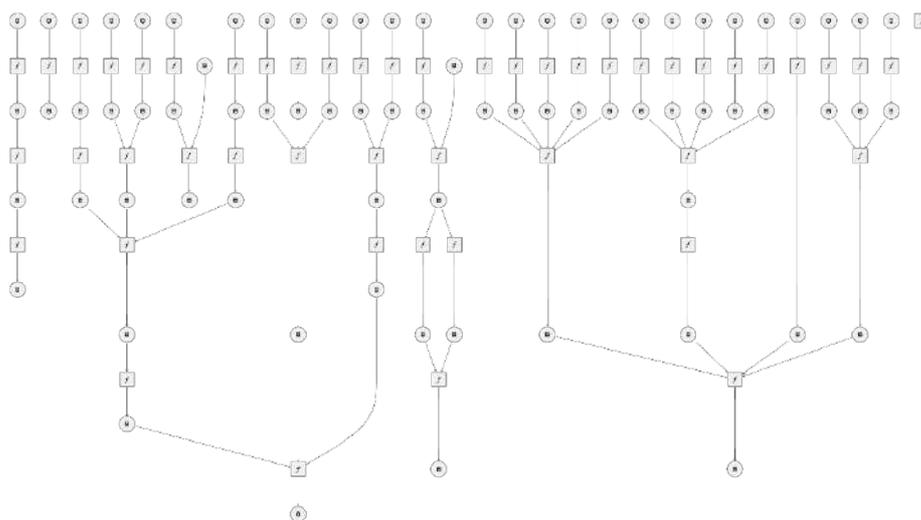


Fonte: Elaborada pelos autores.

Fundamental destacar que não existe a pretensão de, neste artigo, descrever todo o sistema, discorrer sobre todas as funções de transformação e detalhar todas as tabelas geradas, dada a complexidade do trabalho desenvolvido. A Figura 3, no

entanto, mostra, de forma simplificada, os *pipelines*, nós (representados por quadriláteros com a letra “F” em seu interior) e tabelas (imagens circulares) que constituem o sistema *Pool of Features*.

Figura 3 – Representação esquemática de todos os *pipelines* que constituem o sistema *Pool of Features*



Fonte: Elaborada pelos autores.

4 USOS DO SISTEMA *POOL OF FEATURES*

Nesta seção, serão listadas algumas soluções elaboradas com dados do sistema *Pool of Features*. A lista a seguir não é exaustiva e evidencia apenas os painéis estatísticos que estão prontos para publicação e utilização pela Auditoria-Fiscal do Trabalho. Já os painéis a seguir apresentados utilizam o software *Qlik Sense Enterprise* (QLIK, 2021).

Os dados sobre o desempenho da informalidade são fundamentais para melhorar a atuação no combate à precarização das relações de trabalho, privilegiando, assim, o respeito aos direitos trabalhistas mínimos e protegendo contra a incidência do *dumping social*.

Nessa linha, conhecer onde se concentram os maiores indícios de atividades desenvolvidas com base na informalidade dos trabalhadores é fundamental para atuar de forma preventiva e eficaz.

Dessa maneira, os dados presentes na amostra de dados da PNAD Contínua que integram a *Pool of Features* são tratados e, então, geram dados totalizados por diversos níveis de agregação territorial do País, como por estado e por região geográfica. Abaixo, segue a figura da tela inicial do sistema de informalidade:

Figura 4 – Tela inicial do sistema de informalidade no Brasil



Fonte: Elaborado pelos autores.

Além disso, considerando a forma de apresentação dos dados da PNAD Contínua, abertos até o nível de divisão da CNAE, os dados foram sistematizados para agrupar até esse nível, partindo do nível anterior - seção da CNAE. A seguir, na figura 5, está a tela com as opções de filtros permitidos para manipulação desses dados:

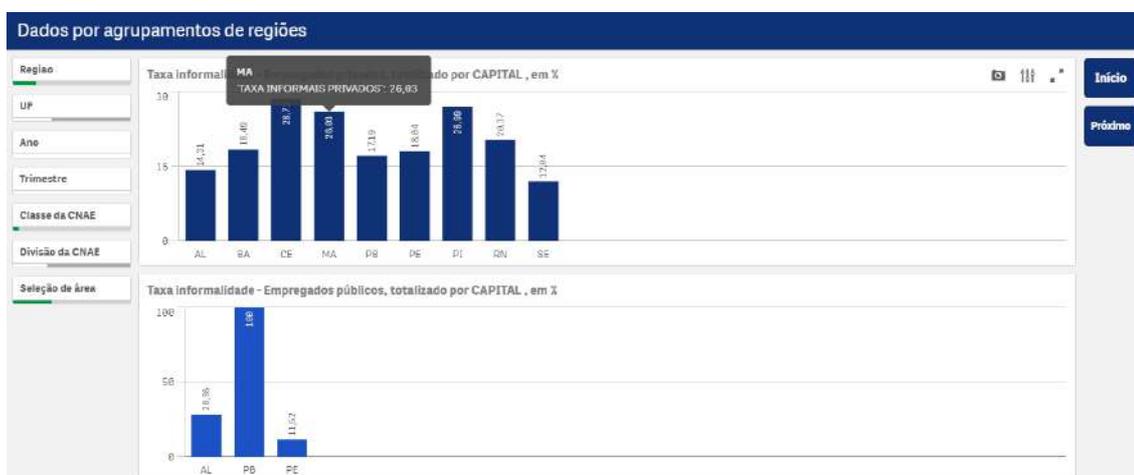
Figura 5 – Tela do sistema de informalidade no Brasil, detalhando-se filtros e gráficos disponíveis



Fonte: Elaborado pelos autores.

O usuário pode trabalhar com os dados de forma interativa para fazer filtros que sejam o mais próximo possível da sua necessidade, atendendo a características peculiares ou específicas do caso concreto. Por exemplo, na Figura 6, abaixo, executa-se através do agrupamento dos dados apenas da região Nordeste para o nível de Capitais e dentro da Seção da CNAE C - Indústrias da Transformação, permitindo-se aferir rapidamente o comportamento da taxa de informalidade observada com tais critérios:

Figura 6 – Exemplo de visualização do sistema de informalidade na região Nordeste com dados agrupados por Capitais



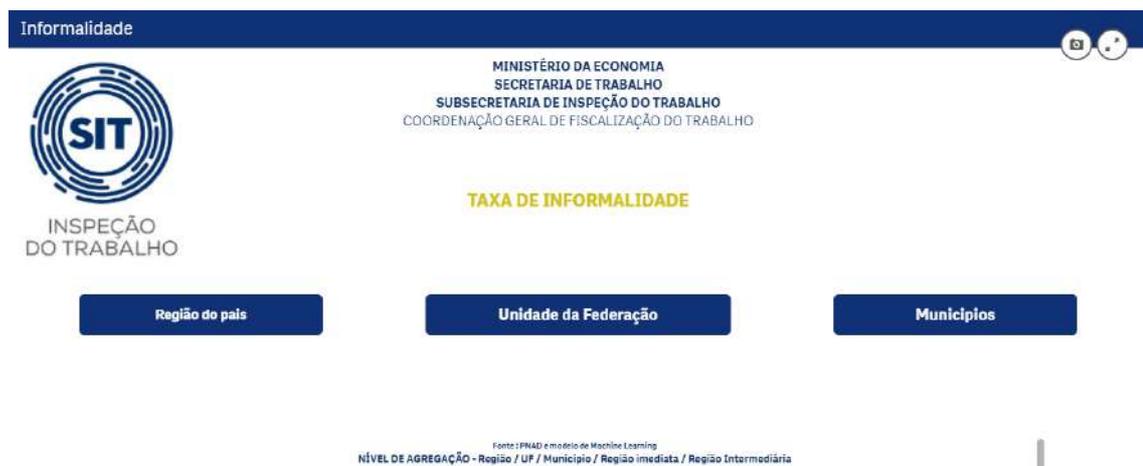
Fonte: Elaborado pelos autores.

Nota-se que a informalidade, neste caso, é apresentada em dois níveis de dados: empregados privados e empregados públicos. Nesse último recorte, apenas em três estados existem dados na PNAD com valores, ao contrário da taxa de informalidade de empregados privados, que é verificada em todos os nove estados da região selecionada.

Em outra frente, foi construído o modelo de *Machine Learning* (aprendizagem de máquina) para calcular as taxas de informalidade em outros níveis de dados não abordados na PNAD Contínua. Assim, para os municípios, regiões imediatas e regiões intermediárias não abrangidas na pesquisa PNAD, o modelo de *Machine Learning* calcula as taxas de informalidade com base nas *features* anteriormente citadas.

Abaixo na Figura 7, está a tela inicial do sistema compilado dos dados.

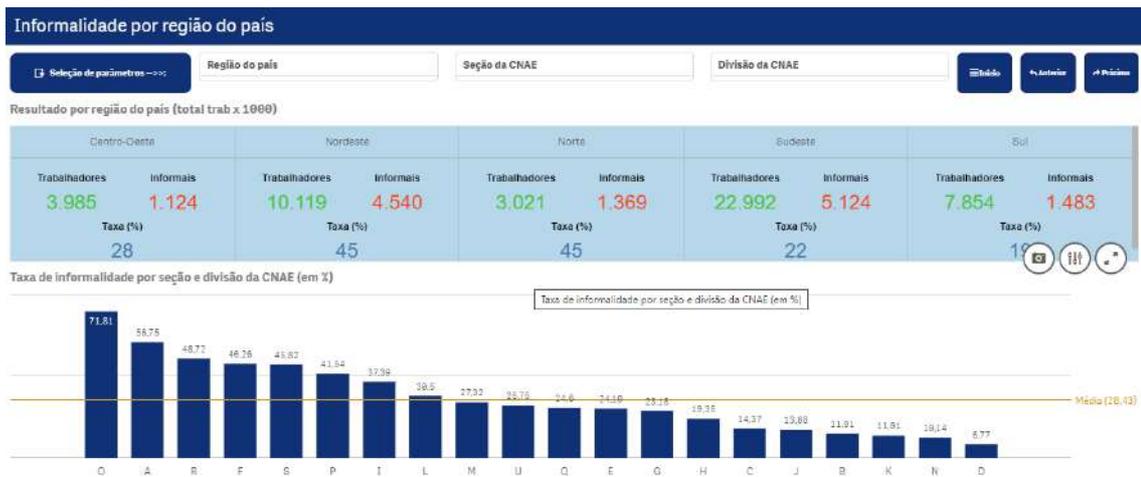
Figura 7 – Tela inicial do painel de informalidade com dados gerados pelos algoritmos de inteligência artificial



Fonte: Elaborado pelos autores.

Em seguida, detalhe da tela que aparece na Figura 8 de consulta da informalidade por região do País, sem utilização de filtros adicionais:

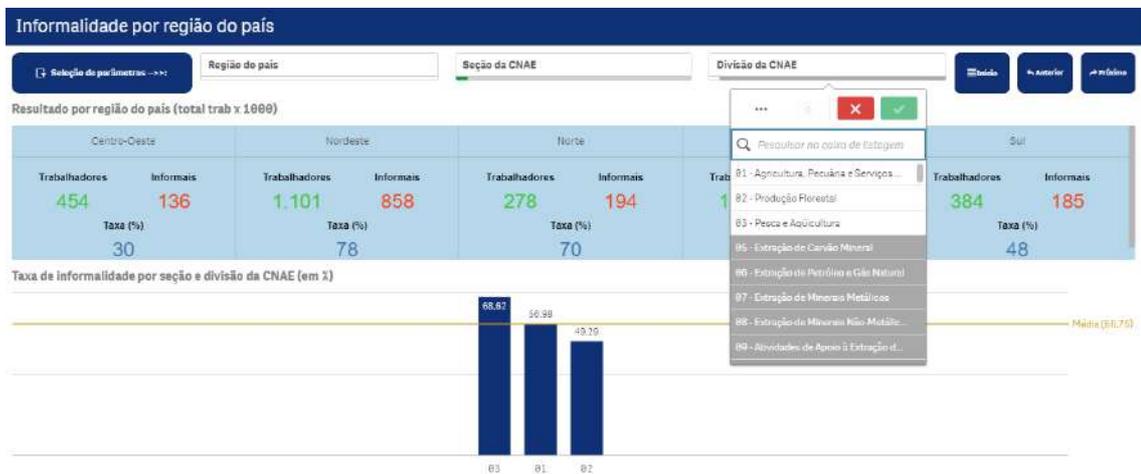
Figura 8 – Informalidade por região do País



Fonte: Elaborado pelos autores.

Nota-se que os dados revelam um alto índice de informalidade nas regiões Norte e Nordeste quando comparadas às demais regiões do País. Outra possibilidade, caso se queira conhecer o comportamento da informalidade dentro da Seção A – Agricultura, Pecuária, Produção Florestal, é filtrar a seção A da CNAE, cujas divisões serão detalhadas na parte inferior da tela, conforme mostrado na Figura 9 a seguir:

Figura 9 – Detalhamento da informalidade na Seção A da CNAE



Fonte: Elaborado pelos autores.

Esta é uma vantagem do uso de ferramentas de *Business Intelligence* – BI para tratamento e consolidação dos dados. É possível obter *insights* e respostas rápidas com pouca manipulação ou esforço, sendo possível combinar e gerar resultados para

tomada de decisões de forma mais assertiva e focada nos resultados apresentados. Passa-se, assim, a ser privilegiada a seleção de empresas ou setores sem o uso de critérios aleatórios, mas sim ancorada em dados numéricos e informações que são, até segunda ordem, confiáveis e oficiais para uso por toda a Administração Pública.

A Inspeção do Trabalho no Brasil vem desenvolvendo sistemas de planejamento com base em algoritmos de inteligência artificial, no intuito de auxiliar no planejamento de ações, principalmente quando se há poucas informações e indícios (TOLEDO; TIMOTEO; BARBOSA, 2020).

Nesse sentido, já foi desenvolvido o sistema capaz de prever a ocorrência de acidentes laborais em determinado empregador para cada CBO (Código Brasileiro de Ocupações) presente em seu quadro funcional (TOLEDO; TIMOTEO; BARBOSA, 2020). Sendo utilizadas, como variáveis para predição, características do empregador - como seu CNAE, o município de sua localização, número de empregados, número de irregularidades previamente detectadas pela Inspeção do Trabalho, número de embargos e interdições já realizados - e características das relações de emprego - idade média dos trabalhadores, tempo médio de trabalho, proporção de mulheres na força de trabalho. Segue, na Figura 10, tela do referido sistema de predição de acidentes laborais com os dados e filtros disponíveis ao usuário:

Figura 10 – Probabilidade de acidentalidade e seus filtros

Maiores probabilidades por UF		Maiores probabilidades por CNAE	
UF	Q	CNAE Subclasse	Q
PE	97,25	Totais	97,25
PR	96,96	4120400 - Construção de edifícios	97,25
SC	96,79	4399103 - Obras de alvenaria	96,96
SP	95,81	4399107 - Montagem e desmontagem de andaimes e out	95,81
MT	94,11	4312600 - Perfurações e sondagens	94,74
RS	93,49	4399199 - Serviços especializados para construção	94,6
MG	93,41	4330402 - Instalação de portas, janelas, tetos, di	94,58
RN	92,54	4321500 - Instalação e manutenção elétrica	94,11
RJ	92,09	4313600 - Obras de urbanização - ruas, praças e ca	93,63
GO	91,94	4320100 - Instalação e manutenção de estruturas	93,77

Dados consolidados - Probabilidade de acidentes por CNPJ													
Ano	Q	UF	Q	CNAE	Q	CNPJ	Q	Raça Social	Q	Município	Q	Probabilidade acidente (%)	Q
2018		PE		4120400		[redacted]		[redacted]		Oitinda		97,25	
2018		PR		4399103		[redacted]		[redacted]		Curitiba		96,96	
2018		SC		4120400		[redacted]		[redacted]		Indaial		96,29	
2018		SP		4399102		[redacted]		[redacted]		Piracicaba		95,81	
2018		SP		4312600		[redacted]		[redacted]		Tupã		94,74	
2018		SP		4399199		[redacted]		[redacted]		Ribeirão Preto		94,6	
2018		SP		4330402		[redacted]		[redacted]		São Paulo		94,58	

Fonte: Elaborado pelos autores.

Na tela acima, é possível ao usuário do painel de dados modelar suas consultas de várias formas, utilizando filtros distintos. Assim, pode-se pesquisar por UF (Unidade

Federativa) ou diretamente pela atividade econômica e obter uma lista de empresas que atendam a tais critérios, sendo apresentada a probabilidade de acidentes com aquela configuração desejada.

Estes parâmetros de consulta fornecem, de forma rápida e ágil, um quadro comparativo dos diversos dados utilizados, sendo a probabilidade calculada reflexo das correlações e efeitos obtidos e armazenados nas informações.

Esse modelo precisa ser revisitado e melhorado, focando na obtenção de resultados cada vez mais precisos para a Fiscalização. Para tanto, além da atualização dos dados com informações mais recentes, é fundamental se notar que há variáveis externas aos bancos da Inspeção do Trabalho que podem contribuir para o modelo construído, como dados de população ocupada no local, informação sobre níveis de educação da região etc.

Além da predição da ocorrência de acidentes de trabalho, outros modelos podem auxiliar no planejamento de ações de segurança e saúde no trabalho. É possível usar os dados contidos na *Pool of Features* para criar modelos mais específicos, tais como para prever a probabilidade de ocorrências de acidentes graves, fatais ou adoecimentos laborais.

Para cada um desses modelos, pode-se definir novas variáveis para usar como parâmetros dos modelos, ou ainda rebalancear o efeito delas dentro do modelo gerado para se obter um resultado mais apurado. Com base nisso, as ações de planejamento podem ser auxiliadas com dados que tornem mais eficaz a seleção de empresas, setores ou períodos do ano para a realização de incursões fiscais.

6 CONSIDERAÇÕES FINAIS

A transformação digital por que têm passado governos e sociedade em todo o mundo não pode ser olvidada pela Auditoria-Fiscal do Trabalho brasileira. Ao contrário, o aumento da disponibilidade de informações em formato digital e da capacidade de armazenamento e processamento de dados tem estimulado o desenvolvimento de novos sistemas para auxílio ao planejamento de ações que permitem alcançar uma maior efetividade na prestação do serviço público.



Importante ressaltar que essa é uma necessária mudança de paradigma e significa um passo importante na direção de embasar o planejamento das ações fiscais a partir da inteligência prévia extraída das informações disponíveis. Não se pode conceber a ausência da Inspeção do Trabalho de uma posição de liderança e protagonismo, dentro do serviço público, na utilização de dados que permitam um salto qualitativo na prestação do serviço e no melhor uso dos recursos disponíveis. A força motriz que permitirá à Auditoria-Fiscal do Trabalho a manutenção da excelência da sua contribuição à sociedade está, indubitavelmente, atrelada a essa mudança estratégica na utilização dos dados e das tecnologias correspondentes.



Nesse contexto, apresenta-se o sistema denominado *Pool of Features* que tem o objetivo de reunir diversas fontes de informação externas à Fiscalização Trabalhista em um único ambiente. Como se tem discutido, o sistema não representa um fim em si mesmo, mas é uma importante fonte de dados para os diversos sistemas de planejamento, avaliação e controle da atuação dos auditores do trabalho. De fato, diversas variáveis importadas pelo sistema em apreço permitem o monitoramento da sociedade em diversos níveis territoriais, possibilitando a escolha, com maior grau de assertividade, de atividades econômicas para fiscalização. Também são possíveis o monitoramento e o controle das ações e de seus impactos nas populações de trabalhadores do País.

É fundamental destacar que o sistema é, inclusive, parte de uma iniciativa maior de uso de aplicações de inteligência artificial pela Fiscalização Trabalhista brasileira. Embora já existam iniciativas voltadas a inserir algoritmos de *machine learning* na rotina de fiscalização, verifica-se a necessidade de ampliar as características usadas pelos sistemas, além de integrar e estruturar as rotinas usadas pela equipe de trabalho de fiscalização. Com o sistema proposto, será possível otimizar o trabalho desenvolvido pela Subsecretaria de Inspeção do Trabalho e acelerar as entregas de produtos para utilização do corpo de Auditores-Fiscais do Trabalho.

REFERÊNCIAS

BORGES, G. M.; ERVATTI, L. R.; SILVA, L. Desafios para o IBGE nas estimativas populacionais dos municípios brasileiros: aplicação de distintas metodologias. **Seminário Internacional “Estimaciones y Proyecciones de Población: Metodologías, Innovaciones y Estimación de Grupos Objetivo de Políticas Públicas”**. Rio de Janeiro–RJ–Brasil, p. 35, 2011.

BRASIL. Lei nº 10.593, de 6 de dezembro de 2002. Dispõe sobre a reestruturação da Carreira Auditoria do Tesouro Nacional, que passa a denominar-se Carreira Auditoria da Receita Federal - ARF, e sobre a organização da Carreira Auditoria-Fiscal da Previdência Social e da Carreira Auditoria-Fiscal do Trabalho, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/2002/l10593.htm>. Acesso em: 4 jun. 2021.

BRASIL. Decreto nº 4.552, de 27 de dezembro de 2002. Aprova o Regulamento da Inspeção do Trabalho. **Diário Oficial [da] República Federativa do Brasil**. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/d4552.htm>. Acesso em: 4 jun. 2021.

DAHL, Ø.; STARREN, A. **The future role of big data and machine learning for health and safety inspection efficiency**. EU-OSHA: Bilbao, Spain, 2019.

FALK, S.; RÖMMELE, A.; SILVERMAN, M. **Digital Government**. [S.l.]: Springer, 2017.

FILGUEIRAS, F.; FLÁVIO, C.; PALOTTI, P. **Digital transformation and public service delivery in Brazil**. Latin American Policy, Wiley Online Library, v. 10, n. 2, p. 195–219, 2019.

FLAK, L. S.; MOE, C. E.; SÆBØ, Ø. On the evolution of e-government: The user imperative. In: SPRINGER. **International Conference on Electronic Government**. [S.l.], 2003. p. 139–142.

HAQ, M. U. **Reflections on human development**. Londres: Oxford University Press, 1995.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **O IBGE**. Disponível em: < <https://www.ibge.gov.br/institucional/o-ibge.html> >. Acesso em: 01 jun. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **SIDRA, Banco de tabelas estatísticas**. Disponível em: < <https://sidra.ibge.gov.br/home/pimpfbr/brasil> >. Acesso em: 01 jun. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua**. Disponível em: < <https://www.ibge.gov.br/estatisticas/multidominio/condicoes-de-vida-desigualdade-e-pobreza/17270-pnad-continua.html?=&t=o-que-e> >. Acesso em: 01 jun. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Estatísticas do cadastro central de empresas: 2018**. [S.l.]: IBGE, 2018. Disponível em : < <https://sidra.ibge.gov.br/pesquisa/cempre/quadros/brasil/2018> > Acesso em: 01 jun. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Produto Interno Bruto - PIB**. [S.l.]: IBGE, 2021. Disponível em : < <https://www.ibge.gov.br/explica/pib.php> > Acesso em: 17 set. 2021.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (Ipea). **O Ipea - Quem Somos**. Disponível em: < https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=1226&Itemid=68 >. Acesso em: 01 jun. 2021.

JANOWSKI, T. **Digital government evolution: From transformation to contextualization**. Nova Iorque: Elsevier, 2015.

MCKINNEY, W. et al. Pandas: a foundational python library for data analysis and statistics. **Python for High Performance and Scientific Computing**, Seattle, v. 14, n. 9, p. 1–9, 2011.

MUSTCHIN, S.; LUCIO, M. M. **The evolving nature of labour inspection, enforcement of employment rights and the regulatory reach of the state in Britain**. Journal of Industrial Relations, SAGE Publications Sage UK: London, England, v. 62, n. 5, p. 735–757, 2020.

OLIVEIRA, L. A. P. d.; SIMÕES, C. C. d. S. **O IBGE e as pesquisas populacionais**. Revista Brasileira de Estudos de População. São Paulo: SciELO Brasil, v. 22, n. 2, p. 291–302, 2005.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). **Índice de Desenvolvimento Humano**. Disponível em: < <https://www.br.undp.org/content/brazil/pt/home/idh0/conceitos.html> >. Acesso em: 01 jun. 2021.

QLIK. **QLIK SENSE, Plataforma de análise de dados**. QLIK, 2021 Disponível em: < <https://www.qlik.com/pt-br/products/qlik-sense> >. Acesso em: 01 jun. 2021.

QUANTUNBLACK VISUAL ANALITICS LIMITED. **Kedro**. Quantumblack, 2021. Disponível em: < <https://kedro.readthedocs.io/en/stable/> >. Acesso em: 01 jun. 2021.
ROCHA, Tatiane; SALVAGNI, Julice; NODARI, Cristine Hermann. **Evidências da Segurança do trabalho e o nível de escolaridade dos trabalhadores na**

construção civil. Revista gestão e planejamento. Salvador, BA. Vol. 20, (jan./dez. 2019), p. 328-343, 2019.

ROSSUM, G. V. et al. **Python.** 1991.

SUZUKI, T. **International trends in systems for inspection of labor law violations.** Japan Labor Review, v. 13, n. 4, p. 83–97, 2016.

TOLEDO, J. de M.; TIMOTEO, R. D. A.; BARBOSA, E. da S. **Inteligência artificial para predição de acidentes laborais no brasil e sua aplicação pela inspeção do trabalho.** Revista da Escola Nacional da Inspeção do Trabalho. 4ª edição. Brasília: Escola Nacional da Inspeção do Trabalho, 2020.

VARELLA, M. D.; OLIVEIRA, C. G.; MOESCH, F. **Salto digital nas políticas públicas: oportunidades e desafios.** Revista Brasileira de Políticas Públicas, v. 7, n. 3, p. 560–583, 2017.

ZAHARIA, M. et al. **Apache spark: a unified engine for big data processing.** Communications of the ACM, ACM New York, NY, USA, v. 59, n. 11, p. 56–65, 2016.

EXTERNAL SOURCE DATA TO IMPROVE LABOUR INSPECTION IN BRAZIL

ABSTRACT

The digital transformation that has taken place in the world in recent decades continues to affect all spheres of society, including the different branches and levels of Public Administration. The natural consequences of this process were, on the one hand, increasing availability of data and, on the other, a significant increase in the need to use them to improve the provision of services, whether in the public or private sphere. About the Brazilian Labor Inspection, these changes do not go unnoticed and several initiatives are beginning to take shape. However, in this case, a large amount of data needs to be meticulously organized so that it can support and exert its expected usefulness. In this vein, this study proposes the Pool of Features Project ("Pool of Features", in literal translation) which represents the effort to bring, in a structured way, data from external sources that can contribute to the IT environment of the Labor Inspection. in improving the quality of public service provision carried out by the Labor Inspection, promoting action planning based on fiscal intelligence.

Keywords: digital government. public data. labour inspection planning.