

INTELIGÊNCIA ARTIFICIAL PARA PREDIÇÃO DE ACIDENTES LABORAIS NO BRASIL E SUA APLICAÇÃO PELA INSPEÇÃO DO TRABALHO

Jefferson M. Toledo¹
Robson D. A. Timoteo²
Enadio da Silva Barbosa³

1. Introdução. 2. Acidentes de trabalho no Brasil. 3. Aprendizagem de máquina (machine learning) e a construção de modelos preditivos. 3.1. Aprendizagem supervisionada e não supervisionada. 3.2. Algoritmos de aprendizagem supervisionada. 3.3. Avaliando os modelos construídos. 4. Modelos de predição de acidentes de trabalho no Brasil. 4.1. Escolha das variáveis de entrada. 4.2. Construção e avaliação do modelo. 5. Aplicabilidade da predição de acidentes para a fiscalização trabalhista no Brasil. 6. Discussões e considerações finais.

RESUMO

O uso de ferramentas de inteligência artificial tem crescido muito nos últimos anos, impulsionado pelo aumento da capacidade de processamento computacional e pelo acesso a dados. Neste trabalho, propomos a aplicação de modelos de aprendizagem de máquina – ramo da inteligência artificial que possui interseções com estatística e ciência de dados – para a predição de acidentes de trabalho no Brasil. Apresentaremos as etapas necessárias ao desenvolvimento do modelo, da obtenção dos dados nas bases disponíveis à Inspeção do Trabalho à escolha e avaliação dos algoritmos preditivos. Foi possível otimizar os modelos e obter predições com acurácia da ordem de 80%. Discorreremos, ainda, sobre o uso dos modelos para auxiliar no planejamento das fiscalizações trabalhistas e sobre as possibilidades de melhorias do sistema desenvolvido.

Palavras-chave: acidentes de trabalho. aprendizagem de máquina. modelos preditivos.

1 INTRODUÇÃO

A Organização Internacional do Trabalho (OIT) estima que cerca de seis mil trabalhadores morrem por dia devido a acidentes de trabalho ou doenças laborais⁴.

¹ Auditor-Fiscal do Trabalho. Doutor em física, com ênfase na área de gravitação e cosmologia.

² Auditor-Fiscal do Trabalho. Mestre e doutor em ciência da computação pela Universidade Federal de Pernambuco.

³ Auditor-Fiscal do Trabalho. Pós-graduado em engenharia de software e MBA internacional pelo instituto de empresa - Espanha. Graduado em direito e ciências econômicas.

⁴ ORGANIZAÇÃO INTERNACIONAL DO TRABALHO. **World Statistic**. Disponível em: <<https://>

Isso corresponde a um total de cerca de 2,19 milhões (dois milhões e cento e noventa mil) de trabalhadores que perdem a vida anualmente em decorrência da atividade laboral. Agrega-se a esta estatística o fato de, anualmente, cerca de 340 milhões de acidentes de trabalho ocorrerem em todo o mundo⁵.

Segundo dados do Observatório de Segurança e Saúde no Trabalho, mantido pelo Ministério Público do Trabalho, entre 2012 e 2018 foi verificada uma morte por acidente de trabalho no Brasil a cada três horas e quarenta e cinco minutos, perfazendo um total de 16.455 (dezesesseis mil quatrocentos e quarenta e cinco) óbitos no período⁶. Estima-se que, no mesmo intervalo de tempo, tenham ocorrido 4,5 milhões de acidentes de trabalho no país, 351 (trezentos e cinquenta e um) mil dias de trabalho tenham sido perdidos e que os gastos previdenciários correspondentes aproximam-se da cifra de R\$ 79 bi (setenta e nove bilhões de reais)⁷. Como se nota, os adoecimentos laborais e os acidentes de trabalho constituem grave problema de saúde no Brasil e no mundo.

Segundo a literatura especializada, entretanto, os acidentes de trabalho são provocados por uma conjunção de fatores que poderiam ser prevenidos⁸. Nesse sentido, a legislação pátria cria um arcabouço conferindo condição de direito social dos trabalhadores a redução de riscos inerentes à atividade laboral através da observância a normas de saúde, higiene e segurança, nos termos do art. 7º da Constituição Federal de 1988⁹. Por consequência, compete aos empregadores implementarem um sistema de proteção à segurança dos empregados, segundo as exigências da legislação trabalhista.

Ainda no plano constitucional, a proteção à vida e à saúde é direito de todos e dever do Estado¹⁰, que tem por obrigação a implementação de políticas sociais e

[//www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm](http://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang-en/index.htm)>. Acesso em: 25 maio 2020.

⁵ Ibid.

⁶ MINISTÉRIO PÚBLICO DO TRABALHO. **Observatório Digital de Saúde e Segurança do Trabalho**. 2020. Disponível em: <<https://smartlabbr.org/sst>>. Acesso em: 08 jun. 2020.

⁷ Ibid.

⁸ ALLI, B. O. **Fundamental principles of occupational health and safety**. 2 ed. Genebra: OIT, 2008.

⁹ BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [1988]. Disponível em:

<http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 08 jun. 2020.

¹⁰ Ibid., art. 196.

econômicas com o fito de reduzir os riscos ao adoecimento¹¹. No âmbito infraconstitucional, a Consolidação das Leis do Trabalho - CLT estabelece, em seu capítulo V, regras gerais de saúde e medicina no trabalho¹², complementadas pelas Normas Regulamentadoras.

Nessa seara, compete aos Auditores-Fiscais do Trabalho, nos termos do artigo 11 da Lei n. 10.593, de 6 de dezembro de 2002¹³ e do artigo 18 do Regulamento da Inspeção do Trabalho, aprovado pelo Decreto n. 4.552, de 27 de dezembro de 2002¹⁴, a verificação do cumprimento dos dispositivos legais relacionados à segurança e à saúde dos trabalhadores. No exercício de suas atividades, os Auditores-Fiscais do Trabalho têm a competência de, dentre outras ações, inspecionar os locais de trabalho, o funcionamento de máquinas e a utilização de equipamentos, averiguar e analisar situações com risco potencial de gerar doenças ocupacionais e acidentes do trabalho, notificar as pessoas sujeitas à inspeção do trabalho para o cumprimento de obrigações. Detectadas situações que coloquem em grave e iminente risco a saúde e a integridade física dos trabalhadores, os Auditores devem efetuar a interdição de estabelecimento, setor de serviço, máquina ou equipamento, ou o embargo de obra¹⁵.

É possível, em suma, verificar que a proteção à saúde dos trabalhadores é amplamente resguardada pela legislação e cumpre ao Estado o estabelecimento de política que torne eficaz esse direito. Aos Auditores-Fiscais do Trabalho, compete papel fundamental de, na qualidade de autoridades em segurança e saúde no

¹¹ Ibid.

¹² BRASIL, **Consolidação das Leis do Trabalho**, aprovado pelo Decreto-Lei nº 5.452, de 1º de maio de 1943. Disponível em: <http://planalto.gov.br/ccivil_03/Decreto-Lei/Del5452.htm>. Acesso em: 08 de junho de 2020.

¹³ BRASIL. **Lei nº 10.593, de 06 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. Brasília, DF: Presidência da República, [2002]. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/2002/L10593.htm>. Acesso em: 08 jun. 2020.

¹⁴ BRASIL. **Decreto nº 4.552, de 27 de dezembro de 2002**. Aprova o Regulamento da Inspeção do Trabalho. Brasília, DF: Presidência da República, [2002] Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/D4552.htm>. Acesso em: 08 jun. 2020.

¹⁵ BRASIL. **Instrução Normativa nº 142, de 23 de março de 2018**. Disciplina procedimentos de fiscalização relativos a embargo e interdição para a atuação da Auditoria-Fiscal do Trabalho. Brasília, DF: Presidência da República, [2018] Disponível em: <<http://www.in.gov.br/web/dou/-/instrucao-normativa-n-142-de-23-de-marco-de-2018-7889407>>. Acesso em: 08 jun. 2020.

trabalho, participarem da consecução das políticas de proteção ao meio ambiente de trabalho. É nesse sentido que vimos propor, neste artigo, a aplicação de técnicas de computação e de programação estatística no auxílio à fiscalização trabalhista, de modo a tornar possível que se antecipe a ocorrência de acidentes laborais e que se possa, eficientemente, auxiliar empregadores e trabalhadores no gozo de seus direitos.

Aprendizagem de máquina (mais conhecido pelo termo original em inglês *machine learning*) é a área da inteligência artificial que fornece aos computadores a habilidade de aprender a partir de dados¹⁶. As soluções construídas com base nessa tecnologia, depois de treinadas, são capazes de fazer previsões de novos eventos e melhorar seu desempenho à medida que são utilizadas. A aprendizagem de máquina é área interdisciplinar que combina métodos de ciência da computação, ciência de dados e estatística na construção de modelos preditivos.

Na última década, diversas aplicações construídas com base na aprendizagem de máquina surgiram e são utilizadas em larga escala, embora esta realidade não seja perceptível por boa parte da população¹⁷. Sistemas de recomendação – usados, por exemplo, em plataformas de *streaming* –, detecção de *spam* na caixa de entrada de e-mails e sistemas de detecção de fraude são alguns exemplos de uso de *machine learning* com quais temos diuturno contato.

Como era de se esperar, há diversas iniciativas governamentais no sentido de utilização de plataformas modernas para auxiliar na consecução de atividades de interesse público. Por exemplo, a Controladoria Geral da União (CGU) possui um sistema que utiliza aprendizagem de máquina para a análise automatizada das prestações de conta em transferências voluntárias da União. Já o Tribunal de Contas da União (TCU) utiliza o robô Alice (Análise de Licitações e Editais) para buscar irregularidades em licitações.

Em nosso trabalho, propomos o emprego da aprendizagem de máquina para construção de uma plataforma capaz de prever futura ocorrência de acidentes de trabalho no país. Como mostraremos, foi possível calcular a probabilidade de

¹⁶ ABU-MOSTAFA, Y. S.; MAGDON-ISMAIL, M.; LIN, H.-T. **Learning from data**. Nova Iorque: AMLBook, 2012.

¹⁷ JORDAN, M. I.; MITCHELL, T. M. **Machine learning: Trends, perspectives, and prospects**. Nova Iorque: Science.

ocorrência de um sinistro laboral para cada empregador do Brasil. A utilização da ferramenta beneficiará sobremaneira o país, permitindo que a inspeção do trabalho escolha de forma assertiva as empresas para os procedimentos fiscalizatórios, promovendo um meio ambiente laboral seguro.

Este trabalho está organizado como segue. Na seção 2, iremos mostrar um breve panorama estatístico dos acidentes de trabalho típicos no Brasil, deixando claro a necessidade de fiscalizações com o objetivo de auxiliar na prevenção de sinistros laborais. Em seguida, na seção 3, fazemos uma revisão sobre aprendizagem de máquina, focando em modelos supervisionados, que serão utilizados no restante do artigo. Na seção 4, apresentamos o desenvolvimento de uma aplicação de aprendizagem de máquina capaz de prever acidentes de trabalho com uma acurácia que se aproxima dos 80 % (oitenta por cento). Por sua vez, na seção 5, apresentaremos algumas sugestões do uso prático da referida aplicação na fiscalização do trabalho. Por fim, na seção 6, apresentamos as conclusões.

2 ACIDENTES DE TRABALHO NO BRASIL

A Lei n. 8.213, de 24 de julho de 1991, conceitua acidente de trabalho típico como aquele "que ocorre pelo exercício do trabalho a serviço da empresa [...], provocando lesão corporal ou perturbação funcional que cause a morte ou a perda ou redução, permanente ou temporária, da capacidade para o trabalho"¹⁸. Equiparam-se aos acidentes de trabalho as doenças profissionais, aquelas produzidas ou desencadeadas "pelo exercício do trabalho peculiar a determinada atividade", e as doenças do trabalho, adquiridas ou desencadeadas "em função de condições especiais em que o trabalho é realizado e com ele se relacione diretamente"¹⁹.

Quando da ocorrência de acidente ou de doença ocupacional, o empregador é obrigado a emitir a Comunicação de Acidente de Trabalho - CAT até o primeiro dia

¹⁸ BRASIL. **Lei nº 8.213, de 24 de julho de 1991**. dispõe sobre os planos de benefícios da previdência social e dá outras providências. Brasília, DF: Presidência da República, [1991] Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/18213cons.htm>. Acesso em: 08 jun. 2020.

¹⁹ Ibid.

útil da ocorrência²⁰. Em caso de óbito, a comunicação deve ser realizada de maneira imediata e, caso o empregador não elabore CAT, esta pode ser formalizada pelo acidentado, sua família, médico, sindicato ou por qualquer autoridade pública²¹.

Os dados das CAT são utilizados para realizar as estatísticas oficiais de acidentes e doenças de trabalho no Brasil²². Além disso, as informações das comunicações estão disponíveis nos bancos de dados disponíveis à Inspeção de Trabalho. Por essas razões e como a proposta deste artigo é propor a construção de aplicação que auxilie a Auditoria-Fiscal Trabalhista na consecução de suas atividades, faz-se tecnicamente viável o uso dos dados de CAT no que segue.

Uma primeira etapa necessária à construção de modelos preditivos é o estudo das características estatísticas do problema abordado. Nessa perspectiva, iniciamos fazendo um breve retrato dos acidentes de trabalho típicos no Brasil nos anos de 2017 e 2018.

Em 2018, foram elaboradas 623.786 (seiscentas e vinte e três mil, setecentas e oitenta e seis) CATs²³. Delas, 57,0 % referiam-se a acidentes de trabalho típico, 1,4 % a doenças de laborais, 16,9 % a acidentes de trajeto e 14,7 % tinham tipo ignorado. Esse perfil é observado em dados de anos anteriores. Em 2017, por exemplo, 58,3 % das comunicações foram elaboradas em consequência de acidente típico. Assim, é de se observar que as características e números de acidentes laborais em sentido *lato* variam significativamente em cada uma de suas classes. Adicionando-se o fato de que os acidentes de trajeto têm tido uma instabilidade legislativa em seu tratamento, conclui-se pela necessidade de tratar individualmente esses acidentes, os típicos e as doenças ocupacionais. O presente trabalho, portanto, focará apenas nos acidentes típicos.

Primeiramente, iremos analisar a relação dos acidentes de trabalho com o sexo e a idade dos empregados. Na figura 1, apresentamos a pirâmide etária de acidentes de trabalho típicos no Brasil no biênio 2017-2018. É possível verificar que

²⁰ Ibid., art. 22.

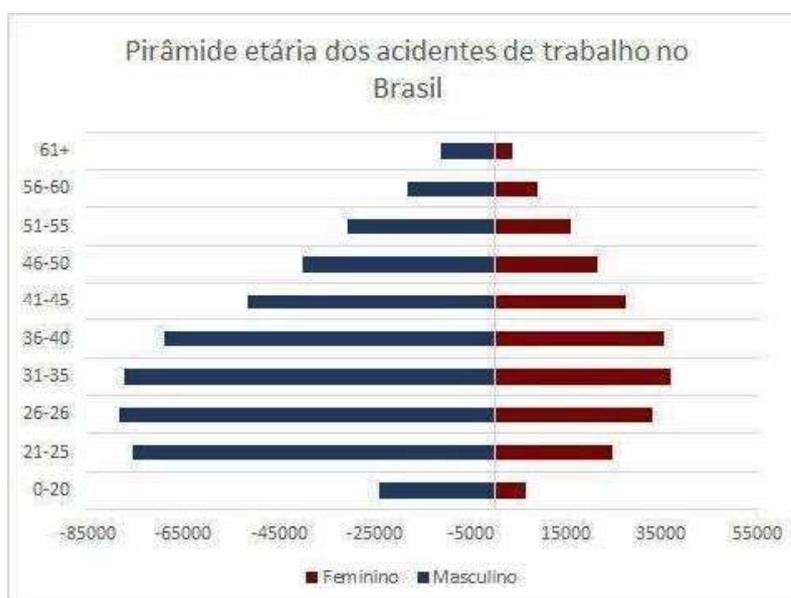
²¹ Ibid.

²² MINISTÉRIO DA FAZENDA. **Anuário Estatístico de Acidentes do Trabalho - AEAT 2017. 2018.** Disponível em: <<http://sa.previdencia.gov.br/site/2018/09/AEAT-2017.pdf>>. Acesso em: 04 jun. 2020.

²³ MINISTÉRIO PÚBLICO DO TRABALHO. **Observatório Digital de Saúde e Segurança do Trabalho.** 2020. Disponível em: <<https://smartlabbr.org/sst>>. Acesso em: 08 jun. 2020.

mais homens que mulheres são acometidos – 67,1 % dos acidentes típicos ocorreram com homens. Além disso, verifica-se concentração de acidentes é maior entre trabalhadores de idades compreendidas entre vinte e quarenta anos. É possível verificar que a pirâmide de acidentes não segue o padrão da população brasileira, que possui mais mulheres que homens, nem a pirâmide etária da população de trabalhadores do país²⁴. Podemos, portanto, concluir que empresas que possuem grandes quantidades de empregados do sexo masculino com idade inferior a quarenta anos tem mais chances de notificar acidentes de trabalho que empresas que empregam mais mulheres com idades mais avançadas.

Figura 1 – Pirâmide etária de acidentes de trabalho típicos no Brasil no biênio 2017-2018.



Fonte: compilação dos autores com base em dados de CATs.

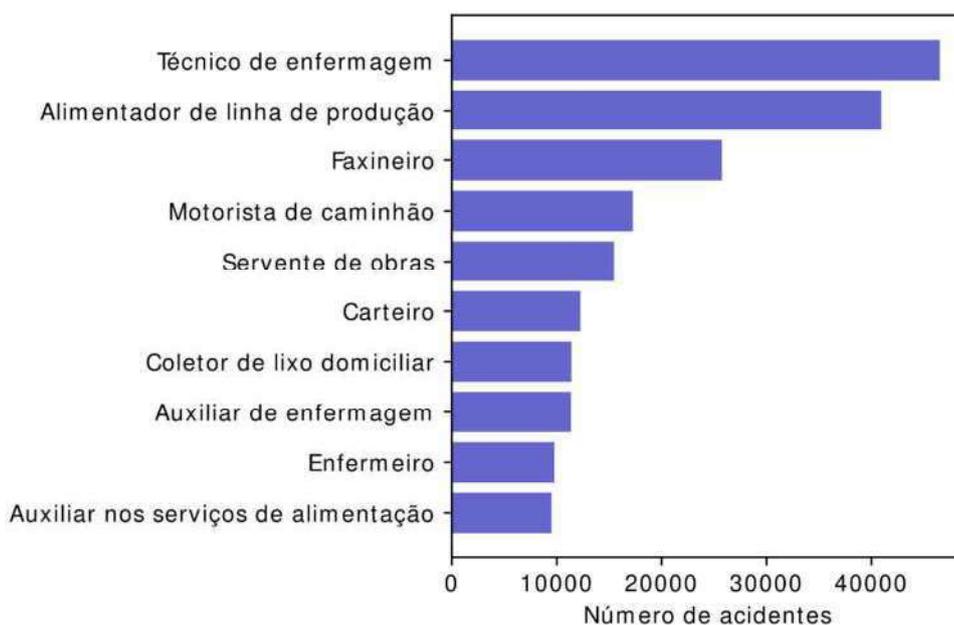
Outros fatores importantes na ocorrência de acidentes laborais são as atividades desempenhadas pela empresa e pelo trabalhador. De fato, as atividades econômicas executadas pelos empregadores apresentam graus de riscos diferentes. Semelhantemente, empregados de uma mesma empresa que realizam diferentes funções em ambientes de trabalho diversos estão expostos a distintas condições de

²⁴ LIMA, A. V. D.; KONRAD, J. A transição demográfica no Brasil e o impacto na previdência social. **Boletim Economia Empírica**, Brasília, DF, v. 1, n. 2, 2020.

segurança em seu meio ambiente laboral.

Na figura 2, apresentamos um gráfico de barras enumerando as dez ocupações que mais acidentaram no biênio 2017-2018. De uma lista de 2.049 ocupações, essas dez atividades corresponderam a cerca de 30 % dos acidentes de trabalho. Pode-se verificar que as atividades de saúde (técnicos e auxiliares de enfermagem e enfermeiro) são aquelas que mais apresentaram CATs por acidente típico no período. Destaque-se, ainda, as atividades de alimentador de linha de produção, faxineiro, motorista de caminhão e servente de obras dentre as ocupações que mais acidentam. Pode-se concluir, conseqüentemente e como se deveria esperar, pela existência de relação entre a função desempenhada pelo obreiro e a acidentalidade laboral.

Figura 2 – Acidentes de trabalho por ocupação para as dez atividades que mais acidentaram no biênio 2017-2018.

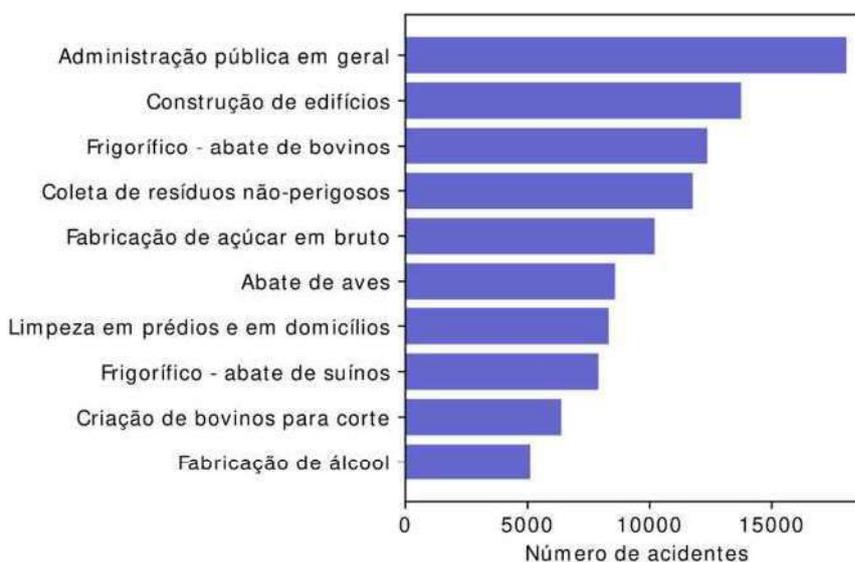


Fonte: compilação dos autores com base em dados de CAT.

Outro fator que guarda relação com acidentes de trabalho no país é a atividade econômica desenvolvida pelo empregador. Na figura 3, representamos um gráfico de barras das dez atividades econômicas que mais provocaram acidentes de trabalho típico em 2017 e 2018. Merece destaque o fato de que o CNAE (Classificação Nacional de Atividades Econômicas) do empregador e o CBO

(Classificação Brasileira de ocupações) do empregado nem sempre estão equitativamente relacionadas nas estatísticas de acidentes laborais. Por exemplo, embora as ocupações de enfermagem levem a emissão da maioria das CATs, as atividades de saúde nem sempre figuram dentre aquelas que mais acidentam. Verifique-se, ainda, que algumas atividades concentram grandes quantitativos de acidentes típicos, como é o caso do cultivo da cana de açúcar e das atividades de frigorífico em geral.

Figura 3 – Acidentes de trabalho por atividade econômica do empregador para as dez atividades que mais acidentaram no biênio 2017-2018.



Fonte: compilação dos autores com base em dados de CAT.

Mais um fator importante na caracterização dos acidentes no Brasil, país continental, é a distribuição geográfica dos sinistros. As unidades federativas diferenciam-se entre si pelas atividades econômicas desenvolvidas e, conseqüentemente, o perfil dos acidentes laborais também se altera. No período de 2017 a 2018, São Paulo foi o estado que mais informou CAT por acidente de trabalho típico, correspondendo a 36,9 % do total do país. A ele, seguiram os estados de Minas Gerais (10,3 %), Rio Grande do Sul (8,8 %), Paraná (8,3 %) e Rio de Janeiro (6,8 %).

As atividades econômicas que mais acidentam também variam entre os

Estados. Excluindo-se a atividade da administração pública, os CNAEs que mais elaboraram CATs de acidentes típicos no Estado de São Paulo foram limpeza em prédios e em domicílios, coleta de resíduos não-perigosos e fabricação de açúcar em bruto, enquanto em Minas Gerais foram os setores de construção de edifícios, cultivo de café e abate de suínos.

Observamos, em suma, que diversas características da empresa e da relação de trabalho podem estar relacionadas à ocorrência de acidentes e podem ser utilizadas para a construção de modelos preditivos. Devido à complexidade do tema abordado, nem sempre é possível concluir pela importância de determinada característica e algumas podem não ter sido aqui citadas. No entanto, os modelos propostos são capazes não só de prever futuros acidentes de trabalho, mas também listar as características que mais foram importantes em sua construção.

3 APRENDIZAGEM DE MÁQUINA (*MACHINE LEARNING*) E A CONSTRUÇÃO DE MODELOS PREDITIVOS

Em 1959, Arthur Samuel definiu a aprendizagem de máquina como sendo “o campo de estudo que dá aos computadores a habilidade de aprender sem serem especificamente programados para tal”^{25,26}. Em outras palavras, a aprendizagem de máquina está relacionada à possibilidade de computadores aprenderem com dados prévios para realizar previsões que melhoram à medida que o sistema vai sendo alimentado com novos dados²⁷.

A aprendizagem de máquina representa uma interseção entre a estatística e a computação, está no núcleo da ciência de dados e é área da inteligência artificial que mais tem se expandido nos últimos anos²⁸. Essa expansão pode ser explicada pela crescente capacidade computacional e de armazenamento de dados e pela

²⁵ Tradução livre do original: “Field of study that gives computers the ability to learn without being explicitly programmed”.

²⁶ SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, IBM, v. 3, n. 3, 1959, p. 212

²⁷ MITCHELL, T. M. et al. **Machine learning**. Nova Iorque: McGraw-hill, 1997.

²⁸ JORDAN, M. I.; MITCHELL, T. M. **Machine learning: Trends, perspectives, and prospects**. Nova Iorque: Science, 2015, p. 255.

disponibilização de informações por meio da internet²⁹. Diversas empresas, instituições de pesquisa e órgãos governamentais em todo o mundo têm utilizado o aprendizado de máquina para auxiliar no desempenho de suas atividades e no aumento da eficiência de suas ações³⁰.

A tecnologia sobre a qual estamos discorrendo deixou, nos últimos anos, de ser curiosidade científica e passou a ter diversas aplicações práticas^{31,32}, com as quais lidamos todos os dias. Os sistemas de recomendação de produtos que grandes empresas utilizam são construídos com algoritmos de aprendizado de máquina. A detecção de e-mails indesejados (*spam*) é realizada utilizando-se algoritmos de *machine learning*.

Mais recentemente, a aprendizagem profunda (*deep learning*) tem ganhado espaço³³, através da do grande desenvolvimento das redes neurais e a consequente criação de aplicações capazes de realizar reconhecimento de sons e de imagens e do processamento de linguagem natural, como tradutores simultâneos.

Vale, por fim, destacar as palavras de M. Jordan e T. Mitchell que, em publicação recente na prestigiada revista *Science*, afirmaram que a aprendizagem de máquina destaca-se como uma das mais importantes tecnologias do século XXI e, portanto, merece ser estudada e amplamente implementada:

Considerations such as these suggest that machine learning is likely to be one of the most transformative technologies of the 21st century. Although it is impossible to predict the future, it appears essential that society begin now to consider how to maximize its benefits.^{34,35}

²⁹ JORDAN, M. I.; MITCHELL, T. M., op. cit., p. 256.

³⁰ Ibid., p. 255.

³¹ Ibid.

³² KUBAT, M. **An introduction to machine learning**. Nova Iorque: Springer, 2017.

³³ SKANSI, S. **Introduction to Deep Learning: from logical calculus to artificial intelligence**. Nova Iorque: Springer, 2018.

³⁴ Considerações como essas sugerem que é provável que a aprendizagem de máquina seja uma das tecnologias mais transformadoras do século 21. Embora seja impossível prever o futuro, parece essencial que a sociedade comece agora a considerar como maximizar seus benefícios (tradução nossa).

³⁵ JORDAN, M. I.; MITCHELL, T. M., op. cit., p. 260.

3.1 APRENDIZAGEM SUPERVISIONADA E NÃO SUPERVISIONADA

Usualmente, os algoritmos de *machine learning* são agrupados em duas classes: algoritmos de aprendizagem supervisionada e modelos de aprendizagem não supervisionada. Em linhas gerais, nos problemas de aprendizagem supervisionada, pretendemos estimar um resultado (conhecido como variável dependente, *output* ou *target*) com base em uma ou mais variáveis de entrada (também denominadas variáveis independentes, *inputs* ou *features*)³⁶. A aprendizagem supervisionada pode ser dividida em problemas de regressão – nos quais se pretende prever o valor da variável de saída com base nas variáveis de entrada – e problemas de classificação – em que, baseados nas variáveis de entrada, é possível dividir os dados em classes previamente conhecidas.

Os modelos de aprendizagem não supervisionada são largamente utilizados para detecção de anomalias ou fraudes num conjunto de dados ou ainda para realizar a classificação de dados em conjuntos não previamente conhecidos – problemas de *clustering*.

Iremos propor neste trabalho a construção de uma aplicação capaz de prever futura ocorrência de acidentes do trabalho em determinada empresa. Como se percebe, o problema com que estamos lidando pretende realizar uma classificação em conjuntos previamente conhecidos. Dessa forma, nosso problema será solucionado por um algoritmo de classificação, como discorreremos, no gênero de aprendizagem supervisionada.

Importa ressaltar que o modelo construído possui boa capacidade de generalização quando é capaz de realizar previsões para conjuntos de dados diferentes daqueles que foram utilizados para desenvolver o seu treinamento. Dessa forma, costuma-se dividir o conjunto de dados utilizados para construção do modelo em dados de treino, dados de validação e dados de teste.

Como sugerido pelo próprio nome, dados de treino, são aqueles utilizados para a construção e treinamento do algoritmo de previsão. Como os modelos possuem certos parâmetros que precisam ser ajustados para melhorar seu

³⁶ JAMES, G. et al. **An introduction to statistical learning**. Nova Iorque: Springer, 2013.

desempenho, utiliza-se o conjunto da validação para comparar o desempenho dos algoritmos com diversos parâmetros, de modo que se possam selecionar aqueles que apresentaram melhor desempenho. Por fim, após haver encontrado o melhor conjunto de parâmetros, utiliza-se o conjunto de testes para analisar as previsões realizadas. Vale destacar que os conjuntos de treino, de validação e de teste não possuem elementos em comum e, com isso, é possível ter confiança no trabalho desenvolvido.

A prática de particionamento do conjunto de dados é recomendada pela literatura especializada³⁷ e foi aplicada nos desenvolvimentos técnicos descritos neste trabalho.

3.2 ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA

Já que o problema abordado neste trabalho será tratado com algoritmos de aprendizagem supervisionada, iremos abordar, em linhas gerais, como eles funcionam.

As variáveis que compõem as características do modelo e que serão utilizadas para sua construção, *features*, serão representadas por X . Por seu turno, a variável de saída será representada por y . Cada exemplo do conjunto de treino é constituído pelos valores das variáveis independentes e da variável dependente, conjunto (X, y) . O nosso objetivo é realizar previsões para conjuntos em que não conhecemos os valores de y . Cada par (X, y) é denominado observação ou instância do conjunto de dados utilizado.

Assim, podemos resumir os problemas de aprendizagem supervisionada no desafio de descobrir qual função ou algoritmo f é capaz de ser utilizado para, sabedores dos valores das variáveis independentes, X , prever os valores da variável dependente em conjuntos previamente não conhecidos, y' . Em outras palavras, temos como desafio determinar f tal que $y' = f(X)$.

Neste artigo, não iremos aprofundar a matemática necessária à construção dos modelos, mas apresentaremos detalhadamente as etapas de sua construção e

³⁷ BISHOP, C. M. **Pattern recognition and machine learning**. Nova Iorque: Springer, 2006.

avaliação, iniciando pelas ferramentas computacionais empregadas.

Python³⁸ é uma linguagem de programação aberta, de propósito geral e que é atualmente a mais ensinada no mundo e aquela mais utilizada para o desenvolvimento de aplicações de ciência de dados³⁹. Dentre os pacotes disponíveis nessa linguagem, o *Scikit-Learn*⁴⁰ apresenta uma série de modelos de aprendizagem supervisionada bem conhecidos e amplamente testados. Durante as etapas práticas de programação e desenvolvimento técnico que culminaram nesta exposição de resultados, diversos algoritmos de aprendizagem supervisionada foram parametrizados e testados com o problema que abordamos. Podemos citar, por exemplo, os modelos de regressão logística, *K-Nearest Neighbor(KNN)*, *Support Vector Machine*⁴¹, *Random Forest*⁴² e *LightGBM*⁴³. Trataremos dos modelos que trouxeram melhores resultados e que poderão ser implementados pela fiscalização do trabalho.

3.3 AVALIANDO OS MODELOS CONSTRUÍDOS

É fundamental destacar que os modelos construídos precisam ser avaliados para que se possa garantir sua utilização com eficácia e para que se possa buscar sua melhoria contínua⁴⁴. Como já afirmado, não é o objetivo deste artigo detalhar a matemática por trás da construção dos modelos. Logo, iremos resumir brevemente algumas das principais métricas de avaliação de desempenho dos algoritmos para predição de classes binárias – apenas duas classes– que, como veremos, é o tipo de classificação que melhor se adequou ao problema que abordamos.

Nessa senda, a acurácia é a razão entre o número de predições corretas e a quantidade total de registros existentes no conjunto de dados. Num problema de

³⁸ ROSSUM, G. V. et al. Python programming language. *In: USENIX ANNUAL TECHNICAL CONFERENCE*, 2007. v. 41, p. 36.

³⁹ SRINATH, K. Python—the fastest growing programming language. **International Research Journal of Engineering and Technology (IRJET)**, v. 4, n. 12, p. 354–357, 2017.

⁴⁰ PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

⁴¹ VAPNIK, V. N. **The nature of statistical learning theory**. Nova Iorque: Springer-Verlag, 1995.

⁴² BREIMAN, L. **Random forests. Machine learning**. Nova Iorque: Springer, 2001.

⁴³ KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *In: GUYON et al. Advances in neural information processing systems*. Nova Iorque: Curran Associates, 2017.

⁴⁴ BISHOP, C. M., op. cit.

classificação binário – que, como veremos, é o que melhor se adequou ao problema que propomos – a acurácia não deve ser tomada como única métrica de avaliação. Isso porque é possível que o modelo desenvolvido tenha uma boa acurácia e faça previsões confiáveis em apenas uma das classes. Por essa razão, utilizaremos, para comparar os modelos concebidos, a precisão e o *recall*.

Num modelo binário, a variável que desejamos prever (y) vai adquirir dois valores: a predição será positiva ($y = 1$) ou negativa ($y = 0$). Com isso, definimos precisão como a razão entre a quantidade de acertos realizados na predição da classe positiva e o total de casos previstos como positivos pelo modelo. A precisão é uma boa métrica para se verificar se a solução está realizando grande predição de falso-positivos⁴⁵. Precisão baixa implica a classificação de muitas observações como falso-positivos.

Por seu turno, o *recall* é calculado pela razão entre o total de predições corretas da classe positiva e o total de predições efetuadas nessa classe. Com o *recall*, podemos verificar se o modelo está realizando um grande número de previsões na classe negativa que deveriam ter sido classificadas como positivas⁴⁶. Em outras palavras, essa métrica, quando baixa, nos informa que há grande número de falso-negativos sendo previstos.

4 MODELOS DE PREDIÇÃO DE ACIDENTES DE TRABALHO NO BRASIL

Como já mencionado, o objetivo do presente estudo é propor a utilização de aprendizado de máquina para a predição da ocorrência de acidentes de trabalho típicos no Brasil.

Vale destacar que um dos desafios da construção do modelo é determinar qual a variável de saída (y) que se pretende obter com o modelo. Nesse intento, vale destacar que, no ano de 2017, 115.276 (cento e quinze mil, duzentos e setenta e seis) empregadores declararam CAT por acidente de trabalho típico no país. Desses, em 76.660 (setenta e seis mil, seiscentos e sessenta) apenas um acidente foi registrado, o que equivale a 66,4 % do total. Situação semelhante ocorreu em 2018,

⁴⁵ BISHOP, C. M., op. cit.

⁴⁶ JAMES, G. et al., op. cit.

quando 122.315 (cento e vinte e dois mil, trezentos e quinze) empregadores apresentaram CAT por acidentes típicos e em 65,7 % apenas um acidente foi registrado.

Percebe-se, assim, que a maioria dos empregadores que informam acidentes de trabalho declara a ocorrência de apenas um infortúnio. Além disso, diversas tentativas foram realizadas no sentido de testar modelos que pudessem prever o número de acidentes que a que ocorrerão em determinada empresa. Os resultados, no entanto, não foram bons quando se tentava diferenciar o número de acidentes. Por esses motivos, o objetivo do nosso trabalho será prever se determinada empresa irá acidentar ($y = 1$) ou não ($y = 0$) e com qual probabilidade o acidente ocorrerá. Como veremos, trabalhando com essa variável de saída, pudemos produzir bons modelos preditivos.

Outro ponto importante do modelo é a determinação do nível de agregação em que serão realizadas as predições. Os diversos testes práticos desenvolvidos apontaram como uma boa instância de predição o par CNPJ-CBO. Isso significa que as predições são realizadas para cada tipo de ocupação dentro do empregador. Com isso, uma mesma empresa pode apresentar probabilidades de acidentes diferentes para empregados que desempenham distintas funções.

4.1 ESCOLHA DAS VARIÁVEIS DE ENTRADA

Como passo inicial para a construção do algoritmo preditivo de acidentes laborais, precisamos determinar que variáveis da relação de trabalho são importantes para a ocorrência do infortúnio. Em outras palavras, devemos estabelecer que características do empregador e do trabalhador são fundamentais para que possamos realizar predições úteis. Constataremos que os modelos também são capazes de avaliar dentre as variáveis de entrada aquelas que mais influenciaram nas previsões.

Como verificamos na seção 2, a distribuição etária de acidentes laborais no Brasil não segue padrão da população do país. Dessa maneira, é possível verificar que a maior concentração de empregados masculinos e a prevalência na faixa etária compreendida entre vinte e quarenta anos aumentam a probabilidade de ocorrência

de acidente de trabalho e, por conseguinte, essas duas características serão utilizadas para construção de modelos de predição.

Outros aspectos importantes e que caracterizam os acidentes laborais são a atividade econômica do empregador e a atividade desempenhada pelo acidentado, como discorrido na seção 2. Ainda nesse sentido, as atividades econômicas são distintas ao longo do território nacional e, por consequência, o comportamento dos acidentes são diferentes em dada região. Outra característica importante que será utilizada para a construção do modelo é, portanto, o município do estabelecimento em que o empregado labora.

O histórico de fiscalizações já realizadas também serviu como base para construção dos modelos. Por experiência, sabemos que empresas que apresentam muitas irregularidades detectadas e muitos itens interditados ou embargado por Auditores tendem a possuir sistemas frágeis de proteção à saúde dos trabalhadores. A vivência nas atividades de fiscalização também corrobora com esse entendimento.

Em suma, na construção dos modelo proposto, usamos como variáveis de entrada o CBO dos empregados, o CNAE nas empresas, o município, o número de empregados, a média de idade dos trabalhadores e as comas, por CNAE e município, de todas irregularidades detectadas pelos Auditores-Fiscais do Trabalho, das irregularidades de jornada e do total de itens interditados ou embargados. Como veremos, além alguns modelos também permitem a avaliar quais variáveis foram mais importantes para as decisões tomadas pelo algoritmo e, com isso, seremos capazes de melhorar o trabalho.

4.2 CONSTRUÇÃO E AVALIAÇÃO DO MODELO

Selecionadas as variáveis de entrada e as classes que se pretende prever, faz-se importante buscar corretamente nos bancos de dados disponíveis à inspeção do trabalho as informações necessárias à construção do modelo.

Como já informado, os dados de acidentes foram buscados dos dados de CAT, incluindo aquelas não elaboradas pelos empregadores, desde que tenha havido o reconhecimento de acidente típico como motivo. As informações de CNAE e município estão presentes nas informações das empresas advindas da Receita

Federal. Os dados de números de empregados, idade e CBO de trabalhadores são coletados nas informações prestadas na Relação Anual de Informações Sociais (RAIS). As informações relacionadas à atuação da Auditoria-Fiscal do Trabalho são colhidas do Sistema Federal de Inspeção do Trabalho (SFITWEB).

Pode-se verificar que as variáveis de entrada são as mais diversas e possuem tipos de dados e limites de valores diferentes. Adotamos, por exemplo, variáveis categóricas (como município e atividade econômica do empregador) e numéricas (como idade e número de trabalhadores). Dentre as numéricas, a quantidade de obreiros de uma empresa pode variar de um até dezenas de milhares, mas suas idades ficam limitadas a algumas dezenas. Para que as variáveis sejam consideradas de forma correta, há a necessidade de uma etapa de pré-processamento, o que foi realizado utilizando-se as ferramentas existentes no pacote *Scikit-Learn*⁴⁷.

Como informamos, o conjunto de dados foi dividido em conjunto de dados de treino, de validação e de teste e os modelos foram construídos na linguagem de programação Python. Cada modelo possui um conjunto de parâmetros que são fixados de acordo com o melhor desempenho apresentado. Na tabela 1, apresentamos as métricas calculadas para quatro algoritmos que apresentaram bom desempenho para o problema abordado.

Perceba-se que os valores de acurácia, precisão e *recall* foram superiores a 74% para todos os algoritmos. Isso significa que, no conjunto de teste, diferente daquele usado para treinar os modelos, foi possível acertar as previsões em boa parte dos dados, mantendo-se baixos níveis de falso-positivos e falso-negativos.

Tabela 1 – Desempenho dos algoritmos de machine learning testados.

Modelo	Acurácia	Precisão	Recall
Regressão logística	0,75	0,75	0,74
KNN	0,77	0,77	0,76
Random forest	0,80	0,80	0,80

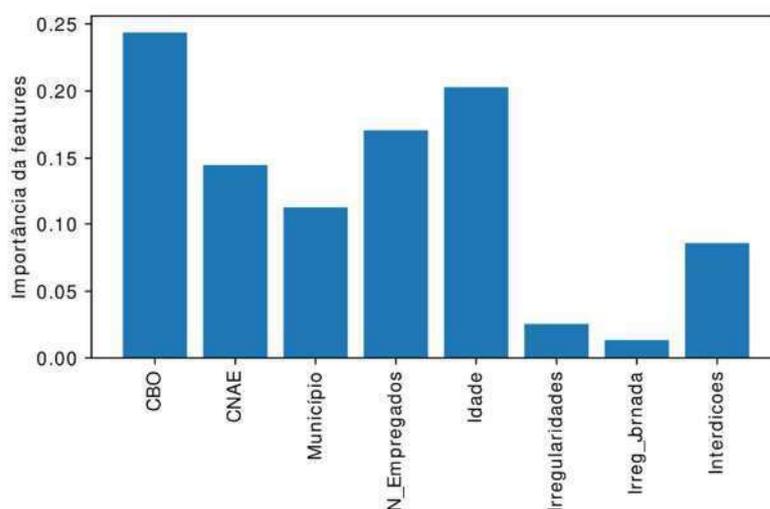
⁴⁷ PEDREGOSA, F. et al., op. cit.

LightGBM	0,74	0,74	0,74
----------	------	------	------

Fonte: métricas de avaliação calculadas para os modelos testados pelos autores.

O algoritmo *Random Forest* permite a avaliação da importância das variáveis de entrada nas predições realizadas, o que pode ser observado na Fig. 4. Nota-se que o CBO do trabalhador e sua idade foram os fatores mais importantes nas escolhas feitas pelo modelo. Das verificações feitas pela Inspeção do Trabalho, as informações de empresas que tenham sofrido embargos e/ou interdições desempenharam, também, importante papel no algoritmo.

Figura 4 – Importância das *features* na construção do modelo.



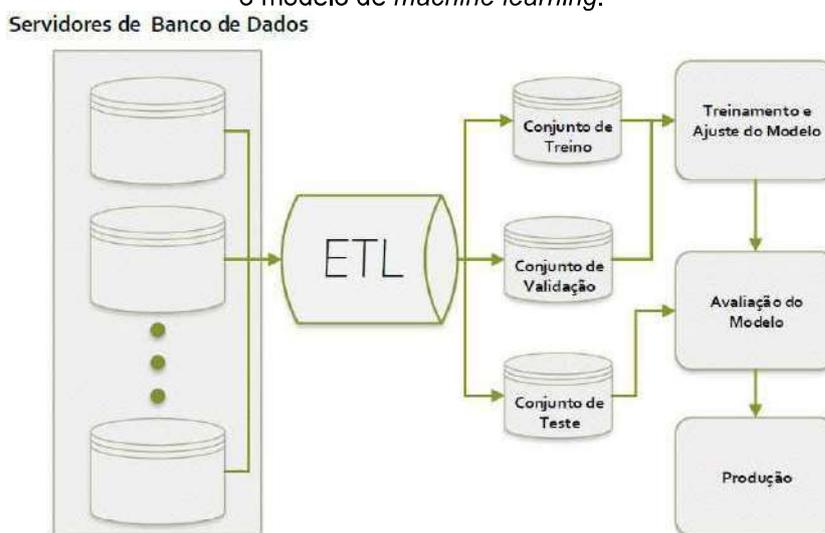
Fonte: compilado pelos autores⁴⁸.

Após a definição das *features* que serão utilizadas pelo modelo, foi criado um sistema com o objetivo de acessar vários bancos de dados e criar os supramencionados conjuntos de treino, validação e teste. Na Figura 5, temos o diagrama básico que ilustra todo o processo para geração do modelo. Inicialmente, temos a extração, transformação e carregamento dos dados (ETL, *Extract Transform Load*). Após esse passo, o modelo é treinado e seus parâmetros ajustados (*tuning*) utilizando os conjuntos de treino e validação. A seguir, o conjunto de teste é utilizado para avaliar a capacidade de generalização do modelo utilizando os melhores

⁴⁸ Importância das variáveis de entrada no modelo *Radom Forest*.

parâmetros encontrados no treinamento. Por fim, após a avaliação, o modelo é colocado em produção e pode realizar a predição de novos dados.

Figura 5 – Diagrama básico ilustrando o processo de extração, transformação e carregamento dos dados (ETL, *Extract Transform Load*) para composição das *features* com objetivo de treinar e avaliar o modelo de *machine learning*.



Fonte: compilado pelos autores.

5 APLICABILIDADE DA PREDIÇÃO DE ACIDENTES PARA A FISCALIZAÇÃO TRABALHISTA NO BRASIL

De acordo com a explanação feita anteriormente, os acidentes no trabalho causam grandes prejuízos, em diversos aspectos, para a sociedade. Eles causam um elevado prejuízo devido à perda de vidas humanas, a incapacitação parcial ou total dos trabalhadores envolvidos e o próprio sofrimento físico associado à sua ocorrência. Além disso, existe o prejuízo financeiro que atinge as empresas e o prejuízo ao ente público que, através de benefícios previdenciários, precisa amparar os trabalhadores ou suas famílias em função do evento acidentário ocorrido.

Para a Inspeção do Trabalho no Brasil, que labora sob a égide do respeito e observação da legislação trabalhista, o que inclui atividades relacionadas à Saúde e Segurança no Trabalho, reveste-se de fundamental importância atuar firmemente na prevenção de acidentes no trabalho. Desta forma, o resultado da utilização de modelos preditivos de *machine learning* pode permitir um avanço significativo na diminuição dos prejuízos humanos e financeiros causados pelos acidentes no

trabalho no país.

A utilização da mencionada tecnologia tem o potencial de expandir o universo de empresas analisadas, gerando um considerável avanço no cumprimento da missão legal dos Auditores do Trabalho. No entanto, faz-se necessário associar a utilização desta tecnologia com uma estratégia que permita a consecução dos melhores e mais rápidos resultados dentro das ações fiscais. Desta forma, sugerimos duas estratégias de uso desse modelo preditivo. Ressalta-se, em tempo, que a contínua utilização desta tecnologia permitirá constatar quais alternativas geram os melhores resultados, além de permitir seu refinamento nesta proposta e o surgimento de outras formas alternativas de utilização.

Esta estratégia pressupõe uma ação concentrada e direcionada na predição e, por conseguinte, prevenção de acidentes do trabalho. O cerne desta proposta consiste na execução do modelo preditivo em um conjunto de empresas de atividade econômicas previamente selecionadas. Desta forma, para operacionalizar esta estratégia, o modelo preditivo seria acionado para analisar uma quantidade específica de CNPJs de um determinado CNAE em um conjunto de regionais escolhidas antecipadamente.

Em seguida, os Auditores Fiscais participantes analisariam as previsões geradas pelo modelo e as empresas seriam notificadas ou fiscalizadas “in loco” a depender da probabilidade das ocorrências de acidentes no trabalho.

Uma segunda estratégia idealizada é introduzir uma funcionalidade no SFITWEB de forma que, ao ser gerada uma nova ordem de serviço (OS), o modelo preditivo possa informar automaticamente quão provável é a ocorrência futura de acidentes naquela empresa. Ato contínuo, esta informação seria analisada pelo Auditor Fiscal do Trabalho responsável pela OS e deliberada com o chefe da Inspeção do Trabalho da respectiva regional. De forma a decidir sobre a ampliação da fiscalização para atender questões específicas ligadas a Acidentes no trabalho ou a criação de uma nova OS direcionada exclusivamente a fiscalizar tal achado.

Observa-se de pronto, o caráter preventivo de tais alternativas, numa tentativa de mitigar os prejuízos anteriormente descritos e tornar o ambiente de trabalho menos nocivo à saúde dos trabalhadores. Acrescenta-se que, uma alternativa para a utilização dos dados preditivos de acidentes no trabalho seria no intuito de

subsidiar o planejamento anual da fiscalização, servindo para gerar a priorização de ações fiscais exclusivamente focados no enfrentamento das questões associadas aos acidentes do trabalho.

6 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Neste trabalho, propusemos a criação de um modelo de aprendizagem de máquina para a predição de acidentes de trabalho no Brasil e verificamos que foi possível obter acurácia da ordem de 80%, realizando-se previsões com baixo número de falso-positivos e falso-negativos. São resultados animadores que permitem à Auditoria-Fiscal do Trabalho se apropriar de avançadas técnicas para trabalhar de maneira preventiva, adiantando-se à ocorrência de sinistros laborais.

Em função da novidade da utilização desta tecnologia, entende-se que uma abordagem inicial possa ser a de implementar esta solução através de projetos-piloto em regionais escolhidas previamente. De fato, como esta tecnologia tem a capacidade de trazer resultados importantes para a fiscalização e para a sociedade, observa-se que a implementação prática deve ser cercada de prudência e de acompanhamento próximo para validar não só a utilização da tecnologia como encontrar a melhor forma de aplicá-la.

Além disso, dado o pioneirismo da iniciativa, o seu êxito pode abrir portas para o desenvolvimento de outros modelos que possam auxiliar outras áreas da fiscalização. Faz-se necessário evidenciar que este é o início de um processo de mudança, onde grandes quantidades de informações prévias podem subsidiar uma melhor gestão dos recursos, uma maior eficiência da fiscalização e um melhor resultado para a sociedade.

Vale salientar também que esta ferramenta tem o caráter de auxiliar o labor do Auditor Fiscal do Trabalho sem, no entanto, jamais ter a pretensão de substituí-lo ou torná-lo obsoleto. A realidade, que já está amplamente estampada na sociedade moderna, é que a enorme quantidade de dados disponíveis deve ser utilizada com o objetivo de agregar e qualificar o conhecimento prévio existente sobre a fiscalização das empresas e sobre o ambiente de trabalho enfrentado pelos trabalhadores no Brasil.

A implementação desta ferramenta não será um ponto final. Será uma porta que se abre no árduo caminho trilhado pela fiscalização trabalhista em direção a lograr uma mais salutar convivência e harmonia dos atores envolvidos com as leis trabalhistas.

Neste diapasão, um pressuposto fundamental deste tipo de tecnologia é sua melhoria constante. O modelo proposto irá sendo melhorado e refinado par-e-passo com o seu uso. Assim como o conhecimento desta tecnologia e as infindáveis possibilidades de utilização para a melhoria do serviço pública na seara circunscrita pela atuação da legislação trabalhista. Desta forma, Aperfeiçoamentos certamente serão necessários e a participação e colaboração de todos será imprescindível.

REFERÊNCIAS

ABU-MOSTAFA, Y. S.; MAGDON-ISMAIL, M.; LIN, H.-T. **Learning from data**. Nova Iorque: AMLBook, 2012.

ALLI, B. O. **Fundamental principles of occupational health and safety**. 2 ed. Genebra: OIT, 2008.

BISHOP, C. M. **Pattern recognition and machine learning**. Nova Iorque: Springer, 2006.

BRASIL, **Consolidação das Leis do Trabalho**, aprovado pelo Decreto-Lei nº 5.452, de 1º de maio de 1943. Disponível em: <http://planalto.gov.br/ccivil_03/Decreto-Lei/Del5452.htm>. Acesso em: 08 de junho de 2020.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [1988]. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 08 jun. 2020.

BRASIL. **Lei nº 8.213, de 24 de julho de 1991**. dispõe sobre os planos de benefícios da previdência social e dá outras providências. Brasília, DF: Presidência da República, [1991] Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8213cons.htm>. Acesso em: 08 jun. 2020.

BRASIL. **Decreto nº 4.552, de 27 de dezembro de 2002**. Aprova o Regulamento da Inspeção do Trabalho. Brasília, DF: Presidência da República, [2002] Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/D4552.htm>. Acesso em: 08 jun. 2020.

BRASIL. **Lei nº 10.593, de 06 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. Brasília, DF: Presidência da República, [2002]. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/2002/L10593.htm>. Acesso em: 08 jun. 2020.

BRASIL. **Instrução Normativa nº 142, de 23 de março de 2018.** Disciplina procedimentos de fiscalização relativos a embargo e interdição para a atuação da Auditoria-Fiscal do Trabalho. Brasília, DF: Presidência da República, [2018] Disponível em: <<http://www.in.gov.br/web/dou/-/instrucao-normativa-n-142-de-23-de-marco-de-2018-7889407>>. Acesso em: 08 jun. 2020.

BREIMAN, L. **Random forests. Machine learning.** Nova Iorque: Springer, 2001.

MINISTÉRIO DA FAZENDA. **Anuário Estatístico de Acidentes do Trabalho - AEAT 2016. 2017.** Disponível em: <<http://sa.previdencia.gov.br/site/2018/04/AEAT-2016.pdf>>. Acesso em: 04 jun. 2020.

MINISTÉRIO DA FAZENDA. **Anuário Estatístico de Acidentes do Trabalho - AEAT 2017. 2018.** Disponível em: <<http://sa.previdencia.gov.br/site/2018/09/AEAT-2017.pdf>>. Acesso em: 04 jun. 2020.

JAMES, G. et al. **An introduction to statistical learning.** Nova Iorque: Springer, 2013.

JORDAN, M. I.; MITCHELL, T. M. **Machine learning: Trends, perspectives, and prospects.** Nova Iorque: Science, 2015.

KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *In*: GUYON et al. **Advances in neural information processing systems.** Nova Iorque: Curran Associates, 2017.

KUBAT, M. **An introduction to machine learning.** Nova Iorque: Springer, 2017.

LIMA, A. V. D.; KONRAD, J. A transição demográfica no Brasil e o impacto na previdência social. **Boletim Economia Empírica**, Brasília, DF, v. 1, n. 2, 2020.

MITCHELL, T. M. et al. **Machine learning.** Nova Iorque: McGraw-hill, 1997.

MINISTÉRIO PÚBLICO DO TRABALHO. **Observatório Digital de Saúde e Segurança do Trabalho.** 2020. Disponível em: <<https://smartlabbr.org/sst>>. Acesso em: 08 jun. 2020.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO. **World Statistic**. Disponível em: <https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm>. Acesso em: 25 maio 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

ROSSUM, G. V. et al. Python programming language. *In*: USENIX ANNUAL TECHNICAL CONFERENCE, 2007. v. 41, p. 36.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, IBM, v. 3, n. 3, p. 210–229, 1959.

SKANSI, S. **Introduction to Deep Learning: from logical calculus to artificial intelligence**. Nova Iorque: Springer, 2018.

SRINATH, K. Python—the fastest growing programming language. **International Research Journal of Engineering and Technology (IRJET)**, v. 4, n. 12, p. 354–357, 2017.

VAPNIK., V. N. **The nature of statistical learning theory**. Nova Iorque: Springer-Verlag, 1995.

ARTIFICIAL INTELLIGENCE FOR PREDICTING OCCUPATIONAL ACCIDENTES IN BRAZIL AND ITS APPLICATION BY THE LABOUR INSPECTION

ABSTRACT

The use of artificial intelligence applications has been growing in the past few years, driven by the raising of computational processing and access to databases. In this work, we propose the application of machine learning – a field of artificial intelligence that intersects statistics and data science – for predicting the occurrence of occupational accidents in Brazil. We present the steps to construct the model, from the data extraction from the Labor Inspection databases to the choice and evaluation of the predictive algorithms. We succeed in model optimization and reached accuracy near 80%. We deal, yet, with the uses of the models to assist in planning the labor inspections and with



the possibility of improving the constructed system.

Keywords: occupational accidents. machine learning. predictive models.