

Mercado de Trabalho

conjuntura e análise

ANO 31 | outubro de 2025

80



PROCESSAMENTO DE LINGUAGEM NATURAL PARA AUTOMATIZAR ETAPAS DA ATUALIZAÇÃO DA CLASSIFICAÇÃO BRASILEIRA DE OCUPAÇÕES¹

Aline S. Martins²
Claudio Bonel³
Enádio da S. Barbosa⁴

1 INTRODUÇÃO

A quantidade e diversidade de ocupações existentes no mundo trabalhista é uma importante referência que ilustra quão diversificada e dinâmica é a economia de um país. Dessa forma, faz-se necessária a criação e manutenção de uma classificação de ocupações que possa refletir essas características inerentes a uma determinada economia, de forma que seja possível a geração de análises e estatísticas para melhor subsidiar políticas públicas a fim de lograr um desenvolvimento equilibrado.

No entanto, o mercado de trabalho se mantém em constante transformação devido a fatores tecnológicos ou mesmo estruturais. Essa contínua transformação provoca a extinção, adaptação e criação de novas ocupações. Isso impõe o grande desafio de manter essa classificação atualizada e aderente à conformação atual da economia, o que significa incluir e alterar ocupações da classificação. Significa também excluir aquelas ocupações que deixam de ser demandadas e não fazem mais parte da vida produtiva da economia nacional.

O processamento de linguagem natural (PLN), segundo Caseli e Nunes (2024, p. 10), “é um campo de pesquisa que tem como objetivo investigar e propor métodos e sistemas de processamento computacional da linguagem humana”. Conforme os mesmos autores, “Na área da Ciência da Computação, PLN está ligado à área da IA [inteligência artificial] e também está intrinsecamente relacionada à Linguística Computacional” (Caseli e Nunes, 2024, p. 10).

Neste artigo, desenvolvemos a hipótese de que o PLN pode ser usado para automatizar etapas do processo de atualização da CBO. Assim, o estudo propõe a criação de modelos de PLN que possam gerar informações de similaridade entre os dados utilizados pelos analistas do MTE, objetivando a análise desses dados de forma mais eficiente, permitindo-lhes uma decisão mais rápida sobre a inclusão, alteração ou exclusão de títulos ocupacionais da CBO. Nesse contexto, quais as contribuições e limitações da utilização de técnicas de PLN, a partir da técnica de similaridade de cosseno, para a avaliação de demandas de criação de CBO?

1. DOI: <http://dx.doi.org/10.38116/bmt80/nt1>

2. Chefe de divisão na Classificação Brasileira de Ocupações do Ministério do Trabalho e Emprego (CBO/MTE). *E-mail*: aline.martins@trabalho.gov.br.

3. Líder técnico em tecnologias disruptivas Linuxcell a serviço do MTE. *E-mail*: claudio.bonel@trabalho.gov.br.

4. Auditor fiscal do trabalho do MTE. *E-mail*: enadio.barbosa@trabalho.gov.br.

1.1 Motivação

A motivação principal deste trabalho reside no fato de utilizar a potencialidade da IA, mais especificamente do PLN, para automatizar atividades repetitivas do processo de atualização da CBO. Isso, de maneira a permitir que a CBO seja atualizada de forma mais rápida, utilizando mais dados, e que os servidores do MTE que desempenham essa tarefa possam se dedicar a atividades mais analíticas.

1.2 Objetivos

O objetivo principal deste trabalho é usar a IA para tornar mais rápido e eficiente o processo de atualização da CBO, auxiliando a análise humana. Já os objetivos específicos são:

- descrever o processo de recepção de demandas de criação e alteração da CBO;
- analisar as contribuições e limitações do PLN, a partir da técnica de similaridade de cosseno entre as demandas da CBO e as informações de perfil ocupacional e de conhecimento do Quadro Brasileiro de Qualificações (QBQ), os títulos de cargos do eSocial e os títulos de vagas de emprego divulgadas na internet; e
- descrever o novo processo de recepção de demandas de criação e alteração da CBO com o uso da IA.

2 A CBO E A PROBLEMÁTICA DE SUA ATUALIZAÇÃO

A Portaria nº 671, de 8 de novembro de 2021, em seu art. 180, § 2º, dispõe que: “a CBO é utilizada nos registros administrativos, para fins classificatórios, sem efeitos de regulamentação profissional” (Brasil, 2021). Tal processo de regulamentação profissional é de competência do Congresso Nacional, nos termos previstos no art. 22, inciso XVI da Constituição Federal de 1988, que assim determina: “Compete privativamente à União legislar sobre: XVI – organização do sistema nacional de emprego e condições para o exercício de profissões” (Brasil, 1988, art. 22).

Assim, cabe ressaltar que a Divisão da CBO (DCBO) atua diretamente no reconhecimento ocupacional, para fins classificatórios, sem efeitos de regulamentação profissional. A CBO é um sistema de classificação de ocupações que tem o objetivo de retratar as diversas atividades laborais existentes no país, de forma padronizada, para fins de levantamentos estatísticos e usos nos registros administrativos.

No Brasil, a estrutura atual da CBO foi criada em 1997; e sua versão vigente, instituída em 2002, materializa a lista de ocupações e profissões do mercado de trabalho brasileiro, sendo instaurada pela Portaria ministerial nº 397, de 9 de outubro de 2002, do MTE (Brasil, 2002). Ela visa à padronização e categorização das ocupações no mercado de trabalho brasileiro. Essa classificação tem características hierárquicas, organizando as ocupações em grupos, subgrupos, famílias e subfamílias, de acordo com as características comuns das atividades desempenhadas, o que permite seu emprego em diversos contextos, como na coleta, análise e divulgação de informações e estatísticas de emprego.

A CBO permite a instituição de bases estatísticas que auxiliam o governo na formulação e implementação de políticas públicas em geral, na orientação de políticas de emprego e renda e na formulação de currículos educacionais. Ela, ainda, empresta maior visibilidade e valorização às profissões e aos trabalhadores vinculados a estas.

A efetiva regulamentação das profissões no país ocorre por meio de leis do Congresso Nacional. Exemplificativamente, a regulamentação da profissão de historiador ocorreu através da Lei nº 14.038/2020 (Brasil, 2020).

A atualização constante de qualquer classificação de ocupações é uma tarefa complexa e que desafia países do mundo inteiro. A responsabilidade de manter a CBO e fazer as devidas atualizações compete à DCBO, que faz parte da estrutura administrativa do MTE. Ela está vinculada à Secretaria Executiva (SE), à Subsecretaria de Estudos e Estatísticas do Trabalho (SEET) e à Coordenação-Geral de Estudos e Estatísticas do Trabalho (CGEET). O processo de atualização tradicionalmente utilizado exige consultas manuais a bases de dados e possui um tempo de resposta elevado. Dessa forma, é necessária uma revisão desse processo para permitir respostas mais rápidas às mudanças no mercado de trabalho.

3 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, descreveremos as bases teóricas que nortearam este trabalho. Serão apresentados alguns fundamentos e o detalhamento da estrutura da CBO. Além disso, serão introduzidos conceitos sobre PLN e como ele pode auxiliar no processo de atualização da CBO. Da mesma forma, serão apresentadas as bases de dados utilizadas, pois elas se constituem na fonte para a entrada de dados utilizados nos modelos. Essas informações serão detalhadas nas subseções a seguir.

3.1 Estrutura da CBO

A CBO possui estrutura hierárquica conforme se apresenta na figura 1. Essa estruturação permite uma organização lógica das ocupações, assim como sua aderência com as demandas do mercado de trabalho. Importante mencionar que a CBO possui também informações denominadas sinônimos. Isso significa que um título ocupacional pode ter sinônimos associados a ele para melhor refletir a denominação de ocupações em diferentes áreas do território nacional. Para efeito de exemplificação, a mesma figura apresenta a ocupação número 214410 e todo o detalhamento de sua posição na estrutura da CBO, incluindo os sinônimos associados.

FIGURA 1
Estrutura da CBO

Estrutura da CBO

✓ Estrutura Hierárquica.

- ✓ Exemplo: **Engenheiro.**
- ✓ Grande grupo: **2 – Ciência e Artes.**
- ✓ Subgrupo Principal: **21 – Ciências exatas, física e engenharia.**
- ✓ Subgrupo: **214 – Engenharia e Arquitetura.**
- ✓ Família Ocupacional : **2144 – Engenheiro Mecânico.**
- ✓ Ocupação: **214410: Engenheiro Mecânico Automotivo.**
- ✓ **Sinônimos:**

1. Engenheiro de manutenção de carros.
2. Engenheiro mecânico de automóveis.
3. Engenheiro mecânico (veículos).



BRASIL PROJETO ATUALIZAÇÃO AUTOMÁTICA DA CBO

MINISTÉRIO DO
TRABALHO
E EMPREGO

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

Elaboração dos autores.

Obs.: A ilustração não pôde ser padronizada nem revisada em virtude das condições técnicas dos originais (nota do Editorial).

3.2 Processo vigente de atualização da CBO

Desde 2002, a CBO recebe atualização anual centrada em revisar descrições e incorporar títulos e famílias ocupacionais de setores da atividade econômica, além de segmentos do mercado de trabalho, dependendo da necessidade. A revisão e a atualização são conduzidas por comissões tripartites, que envolvem representantes do governo, empregadores e trabalhadores, através da utilização de uma adaptação da metodologia *developing a curriculum* (desenvolvimento de currículo) – Dacum.

Esse processo tem início com a análise de escopo, em que avaliações de alterações no QBQ ou as demandas enviadas ao MTE são triadas e, posteriormente, direcionadas a especialistas para análise. No caso das demandas, elas podem trazer informações relevantes, tais como: principais atividades desempenhadas pelo ocupante do título em questão, escolaridade e área de atuação.⁵

5. Elas podem ser enviadas por entidades da sociedade civil, como empresas e sindicatos, diretamente pelo seguinte endereço eletrônico do MTE: <https://cbo.mte.gov.br/cbosite/pages/solicitantes/Solicitar.jsf?jsessionid=ThVkcGJyXHjiZbwMnYJMjgZrvUdvt0JT16HhITCx.CBO-SLV01:mte-cbo>.

É possível solicitar as seguintes providências:

- inclusão, revisão ou exclusão de títulos ocupacionais existentes;
- inclusão ou exclusão de atividades e recursos de trabalho; e
- revisão de redação da formação, da experiência e das condições gerais do exercício do trabalho.

Importante salientar que, mesmo que a demanda para a criação de um novo título ocupacional seja rejeitada, ela pode ser incluída como sinônimo de uma ocupação existente. Em todo caso, as empresas e organizações podem livremente organizar suas estruturas de cargos, independentemente das denominações consagradas na CBO. No entanto, é importante que exista uma convergência entre elas, haja vista que a CBO busca uma aderência maior com a natureza das ocupações existentes no mercado de trabalho brasileiro, e não com as especificidades dos nomes de cargos ou funções existentes nas empresas.

No processo de revisão da CBO, a equipe técnica do MTE realiza consultas manuais à base de dados do QBQ, a vagas de emprego publicadas na internet, à base de dados do Ministério da Educação (MEC)⁶ e a outras informações na internet que, a depender do caso em análise, sejam úteis para confirmar a necessidade da revisão.

No entanto, as consultas manuais requerem um elevado tempo para sua realização, o que redundaria na utilização intensiva de servidores para realizar essa tarefa. Dessa forma, percebeu-se que essa etapa repetitiva de consultas manuais poderia ser realizada por meio da IA. Além disso, modelos de IA poderiam permitir que outras bases de dados fossem consultadas, ampliando o escopo de informações à disposição dos analistas e permitindo decisões baseadas em uma análise com maior profundidade. Neste último caso, seria alcançado um cenário em que a análise seria realizada em tempo consideravelmente menor e, ao mesmo tempo, com consulta a um número significativamente maior e mais diverso de informações.

3.3 Processamento de linguagem natural

O PLN é uma disciplina do campo da IA dedicada à interação entre computadores e linguagem natural humana. Ele é um campo de estudo interdisciplinar que integra fundamentos da linguística (semântica e contextualização), ciência da computação e *machine learning* (aprendizado de máquina) para possibilitar que sistemas sejam capazes de compreender, interpretar e produzir textos de forma humanizada no idioma escolhido.

Em PLN, existem técnicas para encontrar a similaridade entre palavras, frases, semântica e contexto. Isso permite que aplicações identifiquem relações entre palavras ou frases com diferenças sutis, incluindo avaliar a proximidade entre termos, considerando semântica e contexto, o que auxilia na interpretação do significado ou na identificação de entidades relacionadas. Isso posto, o PLN permite medir a similaridade entre textos, viabilizando soluções que permitam às máquinas um certo nível de compreensão da linguagem humana, seus contextos e suas variações linguísticas.

6. Disponível em: <https://cnct.mec.gov.br/>.

Segundo Caseli e Nunes (2024), para desenvolver uma solução, que tem como objetivo final o resultado de uma similaridade entre textos, são necessários a análise, a coleta de dados e o seu pré-processamento. No pré-processamento ocorrem atividades específicas, conforme detalhado a seguir:

- minúsculas (*lowercasing*): converte os textos em letras minúsculas;
- tokenização com normalização: separa as palavras em *tokens* e remove sinais de pontuação em geral;
- remoção de *stop words*: consiste na remoção de palavras como artigos, conjunções, preposições, verbos de ligação; e
- lematização: converte as palavras para suas formas básicas (lemas), baseando-se em suas classes gramaticais e no seu contexto.

Após a realização do pré-processamento, utiliza-se uma técnica de vetorização,⁷ que consiste em transformar o texto em representações numéricas vetoriais. Para efeito deste trabalho, optou-se pela vetorização densa, pois ela captura melhor as relações semânticas e contextuais entre as palavras, além de ser a mais apropriada para lidar com sinônimos. Por fim, faz-se a execução de modelos de PLN para o cálculo da similaridade semântica. Para efeito deste trabalho, foi utilizada uma técnica conhecida como medida de cosseno, pois ela é menos sensível à frequência de ocorrência das palavras em um *corpus* do que outras medidas de similaridade, como a distância euclidiana.

4 MATERIAIS E MÉTODOS

Esta seção descreve a metodologia criada para sustentar este trabalho. O conhecimento adquirido por meio das pesquisas realizadas, as experimentações com técnicas de *machine learning* e a análise dos resultados de diferentes configurações dos algoritmos utilizados permitiu o refinamento dos modelos e a melhoria dos resultados.

Este trabalho caracteriza-se como uma investigação da utilização de modelos de similaridade de cosseno, buscando obter a similaridade entre textos localizados nas bases de dados QBQ e eSocial e nas vagas de emprego de anúncios publicados na internet. A escolha dessas bases de dados foi feita por serem de interesse dos servidores do MTE, haja vista possuírem informações e características relevantes para a identificação e caracterização de ocupações profissionais. Em cada base de dados, foram selecionadas informações específicas e vinculadas ao propósito da investigação. No QBQ, foram escolhidos o perfil ocupacional e o conhecimento, pois descrevem as tarefas e o respectivo conhecimento para o exercício de todas as ocupações da CBO. Por sua vez, no eSocial, foi utilizada a informação dos nomes dos cargos que todos os empregadores utilizam em suas empresas. Por último, na base de anúncios de emprego, foram selecionadas as informações de ocupações reais que as empresas buscam no mercado de trabalho. Necessário salientar a importância da complementaridade das informações dessas três bases de dados, que permite fornecer um conteúdo com maior profundidade para a análise humana final.

7. Disponível em: <https://huggingface.co/>.

4.1 Novo processo de atualização da CBO

Em função do uso dos modelos de IA, foi necessária a criação de um novo processo de atualização da CBO, que levou em consideração as seguintes premissas:

- causar o menor impacto possível no trabalho dos analistas;
- substituir as atividades repetitivas de consulta e análise de dados;
- criar interfaces amigáveis e fáceis de usar; e
- consolidar os resultados em um único ambiente de visualização.

Dessa forma, para dar seguimento ao trabalho, foi estruturado um procedimento para coleta e tratamento desses dados e execução dos modelos de IA. Esse novo procedimento se insere no atual processo de análise das demandas, substituindo a consulta manual de dados. Após a execução desse novo procedimento, os resultados de similaridade encontrados são ranqueados. Os dez melhores resultados de cada base de dados são então apresentados para os analistas em três telas construídas no *software* PowerBi, sendo uma tela para cada base de dados de comparação.

Sendo assim, o novo processo reestruturado de análise de demandas também tem início com a análise de escopo, na qual as avaliações de alterações no QBQ e as demandas enviadas ao MTE são triadas. Em seguida, estas são enviadas para inclusão em um banco de dados e, a partir deste ponto, tem início o processamento dos modelos de IA. Em seguida, os resultados das similaridades são armazenados em uma base de dados e, automaticamente, disponibilizados para os analistas avaliarem. Dessa maneira, confirma-se que a análise humana é necessária para a decisão final sobre a demanda. Depois disso, os analistas já podem emitir algum parecer sobre ela, o qual pode versar sobre a rejeição, a aceitação, a pausa ou mesmo o pedido de mais informações ao demandante.

5 RESULTADOS E DISCUSSÕES

Para a demonstração dos resultados, foi selecionada uma análise de escopo relativa a uma demanda enviada ao MTE. Importante salientar que, para efeito de preservação do sigilo, as informações de caráter pessoal ou que possam identificar o demandante não serão apresentadas. A demanda denomina-se *analista de ESG* – ocupação de consultor de negócios, em que a sigla ESG representa as palavras *environmental, social, and governance* (ambiental, social e governança). A seguir, encontram-se informações da análise de escopo.

A demanda tem o objetivo de incluir o título ocupacional de consultor de negócios com a descrição de analista de ESG na CBO. Ela traz as seguintes informações sobre esta ocupação:

- o nome proposto do título é analista de ESG;
- as principais atividades desempenhadas por ele são: coletar e analisar dados relacionados às práticas ambientais, sociais e de governança; preparar relatórios ESG para investidores, *stakeholders* e agências reguladoras; avaliar os riscos e as oportunidades associados às práticas ESG da empresa; e fornecer consultoria interna para a equipe de gestão e outros departamentos sobre como melhorar as práticas ESG e alinhá-las com os objetivos estratégicos;

- seus principais recursos de trabalho são a internet e o computador;
- a educação geral e profissional requerida é a graduação; e
- a experiência e a certificação requeridas são de especialista em ESG.

A demanda ainda descreve a ocupação mencionando que o ESG surgiu no mercado financeiro como uma forma de medir o impacto que as ações de sustentabilidade geram nos resultados das empresas. Além disso, esclarece que a sigla surgiu a primeira vez em 2004, em uma iniciativa chamada UN Global Compact, dentro de um grupo de trabalho do Principles for Responsible Investment (Princípios do Investimento Responsável – PRI), rede ligada à Organização das Nações Unidas (ONU) que tem o objetivo de convencer investidores sobre investimentos sustentáveis (IFC, [s. d.], p. 2).

Além disso, adiciona informações complementares de que tais analistas pesquisam, analisam e relatam as diversas tendências. Usando os dados públicos disponíveis ou coletados, eles tentam obter ideias que podem ser usadas para criar estratégias de ação em diferentes empresas. Os analistas são flexíveis, trabalham em vários setores, com diferentes tipos de base de dados, e podem ser solicitados a dedicar um tempo significativo para gerar e entregar relatórios. Os analistas precisam ter diploma de bacharel em matemática, gestão de empresas, *marketing*, ciência da computação ou de campos relacionados. Mestrado em administração de empresas (MBA), *marketing*, matemática, ciência da computação ou em áreas relacionadas podem ser preferidos. Os analistas se destacam por possuírem habilidades distintas de comunicação e pesquisa. Essa função também requer que sejam atentos a detalhes e que tenham iniciativa.

Dessa forma, os servidores encarregados do MTE analisam a demanda e iniciam a fase de coleta de informações relacionadas a essa ocupação. Nesse momento, o processamento dessa demanda já foi realizado pelos modelos de IA, e as informações de similaridade relativas às bases QBQ e eSocial e às vagas publicadas já podem ser consultadas diretamente nas telas do painel da classificação automática da CBO. Para efeito de simplificação, vamos descrever com detalhes somente a consulta da similaridade relativa à base de dados do eSocial.

Na figura 2, visualizamos o resultado, em ordem decrescente, das dez maiores similaridades da demanda em relação ao nome do cargo na base de dados do eSocial.

FIGURA 2
Visualização da similaridade no eSocial

Similaridade QBQ	Similaridade eSocial	Similaridade Vagas			
Demanda	Familia Ocupacional	Cód. CBO - Ocupação	Cargo eSocial	Similaridade eSocial	
10428 - ANALISTA DE ESG	1421 - Gerentes administrativos, financeiros, de riscos e afins	142125 - Analista de compliance	ANALISTA ESG SR	0,95	
10428 - ANALISTA DE ESG	1423 - Gerentes de comercialização, marketing e comunicação	142330 - Analista de negócios	ANALISTA ESG PL	0,95	
10428 - ANALISTA DE ESG	2140 - Engenheiros ambientais e afins	214010 - Tecnólogo em meio ambiente	ANALISTA ESG SR	0,95	
10428 - ANALISTA DE ESG	2511 - Profissionais em pesquisa e análise antropológica sociológica	251120 - Sociólogo	ANALISTA DE ESG	1,00	
10428 - ANALISTA DE ESG	2521 - Administradores	252105 - Administrador	ANALISTA ESG	1,00	
10428 - ANALISTA DE ESG	2524 - Profissionais de recursos humanos	252405 - Analista de recursos humanos	ANALISTA DE ESG	1,00	
10428 - ANALISTA DE ESG	3115 - Técnicos em controle ambiental, utilidades e tratamento de efluentes	311505 - Técnico de controle de meio ambiente	ANALISTA DE ESG	1,00	
10428 - ANALISTA DE ESG	3522 - Agentes da saúde e do meio ambiente	352205 - Agente de defesa ambiental	ANALISTA DE ESG	1,00	
10428 - ANALISTA DE ESG	3522 - Agentes da saúde e do meio ambiente	352205 - Agente de defesa ambiental	ANALISTA ESG	1,00	
10428 - ANALISTA DE ESG	3522 - Agentes da saúde e do meio ambiente	352205 - Agente de defesa ambiental	ANALISTA ESG SR	0,95	

Elaboração dos autores.

Obs.: A ilustração não pôde ser padronizada nem revisada em virtude das condições técnicas dos originais (nota do Editorial).

A primeira coluna, *demandas*, representa o número e a descrição da demanda para criação ou alteração da CBO enviada ao MTE. As colunas *família ocupacional* e *cod. CBO – ocupação* representam o número e a descrição da família ocupacional da CBO, e o código e a descrição da ocupação encontrados pelo modelo de IA, o qual também identificou a respectiva similaridade. A coluna *cargo eSocial* representa o cargo do eSocial que detém a respectiva similaridade com o título da demanda. A coluna *similaridade eSocial* apresenta o resultado do cálculo da similaridade. Para facilitar a visualização, a tela agrupa os resultados pela coluna *família ocupacional*.

Para efeito de análise, leva-se em consideração não somente o valor da similaridade, mas também a frequência das famílias ocupacionais encontradas. Isso significa que famílias com maior frequência podem sinalizar maior proximidade entre o título da demanda e o local em que essa ocupação se encontra na estrutura da CBO. Para permitir flexibilidade na ordenação dos resultados, é possível realizar um clique duplo em cima do nome de cada coluna para que os resultados sejam ordenados pela respectiva coluna. Um primeiro clique duplo fará a ordenação em ordem crescente; um segundo clique duplo fará a ordenação em ordem decrescente.

Em seguida, como ainda não foram implementados nessa aplicação os modelos de similaridade para as bases do MEC,⁸ a da própria CBO⁹ e as internacionais, os servidores do MTE realizam essas consultas manualmente, quando necessário. Nesse momento, analisam-se todos os dados e conclui-se com o parecer sobre a demanda, conforme se segue:

8. Disponível em: <https://cnct.mec.gov.br/>. Acesso em: 10 out. 2024.

9. Disponível em: <http://www.mtecbo.gov.br/cbosite/pages/pesquisas/BuscaPorCodigo.jsf>. Acesso em: 21 out. 2024.

Para efeitos de CBO, o analista de ESG é entendido como um cargo, que pode ter relação com várias ocupações já existentes na CBO, a depender do conjunto de atividades a serem exercidas pelo ocupante do cargo. Feitas estas considerações, informamos que a demanda em questão obteve parecer desfavorável de inclusão como uma nova ocupação, na estrutura da CBO, neste momento. Ressaltamos que este parecer é passível de alteração em consequência de alteração no cenário avaliado.¹⁰

Como demonstrado anteriormente, a IA foi utilizada para verificar um grande volume de informações que, no caso da base de dados do eSocial, chegou próximo dos 50 milhões de registros. Essa verificação envolveu a análise sintática e a verificação do conteúdo semântico das informações textuais envolvidas, de forma a fornecer uma análise rápida, profunda e que foi capaz de substituir a consulta manual antes realizada pelos servidores responsáveis. Esse resultado gerado pelos modelos de IA auxiliou os analistas a concluírem pela não inclusão da demanda exemplificada anteriormente.

Levando-se em conta que esse processo utilizando a IA também é feito para outras bases de dados e será realizado, invariavelmente, para todas as demandas recebidas pelo MTE, podemos vislumbrar, ao longo do tempo, uma considerável redução na duração do processo e um incremento significativo na agilidade para responder às demandas. Pode-se inferir também que, se mais bases de dados forem incluídas para análise dos modelos de IA, menor poderá ser o tempo de análise de cada demanda. Isso permite que os servidores se concentrem em atividades mais analíticas, contribuindo para aprofundar a discussão e a pertinência dessas alterações sofridas pela CBO, além de auxiliar no desafio de tornar essa classificação cada vez mais atualizada com relação ao mercado de trabalho.

Dessa forma, partir da análise dos resultados alcançados, é possível concluir que a lógica que permeou a estratégia de modelos de IA para substituir a consulta manual de bases de dados se mostrou viável e eficiente. O uso de modelos de PLN contribuiu no tratamento de demandas de atualização da CBO, uma vez que foi possível analisar grandes quantidades de dados de maneira rápida e eficiente. Importante frisar que outras bases de dados podem ser adicionadas a esse processo, permitindo aprofundar a análise das demandas e fornecendo aos analistas um cenário mais completo para a elaboração do parecer. Em contrapartida, entende-se que os modelos baseados em linguística computacional possuem limitações e não conseguem atingir o nível de cognição da análise humana, redundando, em alguns casos, em resultados que possuem proximidade ortográfica, mas se situam distantes em contexto.

6 COMENTÁRIOS FINAIS

A IA é uma tecnologia em constante transformação e evolução. Isso permite que seu uso seja expandido em uma crescente gama de situações e contextos. Novas pesquisas são realizadas e novos algoritmos são criados constantemente, o que incrementa sua adaptabilidade e aumenta a precisão dos resultados que podem ser alcançados. Neste projeto, algumas alternativas podem ser consideradas para melhorar os resultados, a fim de permitir que a atualização da CBO seja feita de forma mais

10. Documento não publicado de resposta do MTE à demanda recebida.

eficiente e consiga uma aderência maior à atualidade das transformações ocorridas na economia e no mercado de trabalho. Entre essas alternativas, podem-se considerar as descritas a seguir:

- adicionar as seguintes bases de dados, com seus respectivos modelos e resultados de similaridade: educação (MEC) e classificações internacionais (Organização Internacional do Trabalho – OIT e Standard Occupational Classification – SOC);
- construir um modelo linguístico próprio, padronizado e com semântica e contextos do MTE; e
- construir um modelo de IA que utilize dados para sugerir automaticamente novas ocupações, para que estas funcionem como uma demanda criada pela IA que seguirá o fluxo de análise e inspeção humana.

Em última instância, esta iniciativa possui alternativas para melhorar os resultados e evoluir para uma solução que dê à atualização da CBO um nível maior de sincronia com as mudanças ocorridas no mercado de trabalho e na vida econômica do país e de seus trabalhadores.

Um grande passo nesse sentido é a possibilidade de criar modelos para sugerir novas ocupações independentemente do envio de demandas por entidades externas. Isso significa que os modelos poderiam sugerir novas ocupações com base em informações existentes no mercado de trabalho, sem prescindir, no entanto, da avaliação e análise de sua pertinência e da adequação pelos servidores do MTE.

REFERÊNCIAS

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília: Congresso Nacional, 1988.

BRASIL. Portaria nº 397, de 9 de outubro de 2002. Aprova a Classificação Brasileira de Ocupações – CBO/2002, para uso em todo território nacional e autoriza a sua publicação. **Diário Oficial da União**, 9 out. 2002. Acesso em: 10 jan. 2025.

BRASIL. Lei nº 14.038, de 17 de agosto de 2020. Dispõe sobre a regulamentação da profissão de Historiador e dá outras providências. **Diário Oficial da União**, Brasília, p. 4, 18 ago. 2020.

BRASIL. Portaria nº 671, de 8 de novembro de 2021. Regulamenta disposições relativas à legislação trabalhista, à inspeção do trabalho, às políticas públicas e às relações de trabalho. **Diário Oficial da União**, 11 out. 2021. Acesso em: 10 jan. 2025.

CASELI, H. M.; NUNES, M. G. V. (org.). **Processamento de linguagem natural: conceitos, técnicas e aplicações em português**. 3. ed. São Carlos: BPLN, 2024.

IFC – INTERNATIONAL FINANCE CORPORATION. **Who Cares Wins, 2004-08**: issue brief. Washington: IFC, [s. d.]. Disponível em: <https://documents1.worldbank.org/curated/en/444801491483640669/pdf/113850-BRI-IFC-Breif-whocares-PUBLIC.pdf>.

BIBLIOGRAFIA COMPLEMENTAR

BONEL, C. **Metodologia e engenharia de requisitos para projetos de *business intelligence***. Joinville: Clube de Autores, 2021.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and Tensor- Flow**: concept, tools, and techniques to build intelligent systems. Sebastopol: O'Reilly Media, 2022.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing**: an introduction to natural language processing, computational linguistics, and speech recognition. 3. ed. Stanford: Stanford University, 2024.

KNUTH, D. E. **The T_EX Book**. 15. ed. Boston; São Francisco; Nova York: Addison-Wesley, 1984.

LIMA, D. M.; GONZALES, L. E. F. **Matemática aplicada à informática**. Porto Alegre: Bookman, 2015.

MANNING, C. D.; SCHÜTZE, H. **Foundations of statistical natural language processing**. Cambridge, Estados Unidos: MIT Press, 1999.

MORETTIN, P. A.; SINGER, J. da M. **Estatística e ciência de dados**. 2. ed. Rio de Janeiro: LTC, 2025.

POTE, S. **Machine learning in production**: master the art of delivering robust machine learning solutions with MLOps. Londres: BPB Publications, 2023.

PRESSMAN, R. S.; MAXIM, B. R. **Engenharia de *software***: uma abordagem profissional. 9. ed. Porto Alegre: AMGH, 2021.

SMITH, A.; JONES, B. On the complexity of computing. *In*: SMITH-JONES, A. B. (ed.). **Advances in computer science**. [S. l.]: Publishing Press, 1999. p. 555-566.

YIN, R. K. **Estudo de caso**: planejamento e métodos. 2. ed. Porto Alegre: Bookman, 2001.

ZAVE, P. Classification of research efforts in requirements engineering. **ACM Computing Surveys**, v. 29, n. 4, p. 315-321, 1997.

Missão do Ipea

Qualificar a tomada de decisão do Estado e o debate público.



ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
PLANEJAMENTO
E ORÇAMENTO

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO