

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Kelvin Henrique Vieira Pedroso

**Algoritmos de *Machine-Learning* aplicados à
predição de salários com base em dados de
admissão do Cadastro Geral de Empregados e
Desempregados – CAGED (2022)**

Curitiba
2023

Kelvin Henrique Vieira Pedroso

Algoritmos de *Machine-Learning* aplicados à predição de salários com base em dados de admissão do Cadastro Geral de Empregados e Desempregados – CAGED (2022)

Monografia apresentada ao Programa de Especialização em *Data Science e Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Luiz Eduardo Soares de Oliveira

Curitiba
2023

Algoritmos de *Machine-Learning* aplicados à predição de salários com base em dados de admissão do Cadastro Geral de Empregados e Desempregados – CAGED (2022)

Machine-Learning algorithms applied on wage predictions based on hiring data of General Register of Employees and Unemployed - CAGED (2022)

Kelvin H. V. Pedroso¹, Luiz E. S. de Oliveira²

¹Aluno do programa de Especialização em Data Science & Big Data*

²Professor do Departamento de Estatística - DEST/UFPR[†]

Modelos de *Machine Learning* aplicados a predição de salários podem ser utilizados pelo setor público como uma ferramenta adicional ao combate à sonegação de imposto de renda e pelo setor privado no combate à taxa de *turnover*. O objetivo deste trabalho é explorar modelos de *machine learning* que possibilitem prever salários. Para isso foi utilizado os dados de admissão de 2022 do Cadastro Geral de Empregados e Desempregados – CAGED e treinados modelos em *Python* da biblioteca *Scikit-Learn*. Os estudos apontaram para os atributos de ocupação, grau de instrução e idade do admitido como os mais relevantes para predição salarial. O modelo com menor erro médio foi o *GradientBoostingRegressor* treinado em uma base de dados com atributos mais desagregados. Também fica patente que a distribuição dos salários majoritariamente concentrados em faixas mais baixas fez tender para baixo as predições e portanto, os modelos não desempenharam bem na predição de altas rendas.

Palavras-chave: CAGED, Aprendizagem de Máquina, predição, salários

Machine-Learning models can be applied to wage predictions that can be used by the government as an additional tool to fight against income tax evasion and by the private sectors against increases in turnover rates. The aim of this work is to explore machine-learning models that allow us to predict wages. For this purpose, the 2022 hiring data from the General Registry of Employed and Unemployed - CAGED was used, and models were trained based on Scikit-Learn library from Python Language. The findings indicate that the occupation, educational level and age of the employee are the most relevant attributes for wage prediction. The model with the lowest mean error was the GradientBoostingRegressor trained on a dataset with more disaggregated attributes. It is also evident that the distribution of wages predominantly concentrated in lower ranges caused the predictions to tend downwards, and therefore, the models did not perform well in predicting high incomes.

Keywords: CAGED, Machine Learning, prediction, wages

1. Introdução

Atualmente, modelos de *Machine Learning* estão presentes nas mais diversas atividades humanas e são utilizados para planejar, produzir, vender e se relacionar com o público [4]. Os algoritmos permitem identificar padrões, tendências e inconsistências, como, por exemplo, em casos de fraude financeira. Desta forma, os benefícios para as organizações estão na tomada de ações mais rápidas, redução de custos com processos

aprimorados e a utilização de recursos de modos mais eficientes. Modelos de *Machine Learning* aplicados a predição de salários podem ser utilizados pelo setor público e privado. Para o setor público sendo uma ferramenta adicional ao combate à sonegação de imposto de renda de pessoas físicas e para o setor privado no combate ao crescimento das taxa de *turnover*.

A sonegação de impostos é um problema na medida em que, por força de lei, as despesas do governo federal estão atreladas à arrecadação. Em cálculo feito pelo Sonegômetro, mantido pelo Sindicato Nacional dos Procuradores da Fazenda Nacional – SINPROFAZ

*vieirapedroso@gmail.com

[†]luiz.oliveira@ufpr.br

estima-se que desde 01 de janeiro de 2023 até final de junho deste mesmo ano mais de R\$ 300 bilhões foram perdidos com sonegação de impostos. Com este valor, segundo o Sonegômetro, seria possível pagar 117 milhões de salários anuais para professores do ensino fundamental (com piso do MEC), pagar mais de 300 milhões de salários-mínimos, distribuir 3,3 bilhões de bolsa família (benefício básico), construir 5,4 milhões de casas populares de 40m², comprar 1,8 milhões de ambulâncias equipadas e pagar 122 mil salários anuais de policiais (PMSP 2ª classe) [8].

O *turnover* é um termo da língua inglesa que no meio empresarial remete à rotatividade dos colaboradores na empresa. Quanto maior a taxa de *turnover* pior tende a ser as consequências para o desenvolvimento dos negócios. O *turnover* portanto, pode se manifestar de forma disfuncional, comprometendo as atividades usuais e estratégicas da empresa que por consequência prejudicam indicadores de produtividade, sustentabilidade, competitividade e rentabilização [1]. Segundo levantamento feito pela consultoria de recrutamento Robert Half, 56% dos executivos *c-level* no Brasil perceberam aumento no *turnover* frente ao período pré-pandemia, essa é a maior taxa entre os países analisados [7]. As principais causas do *turnover* são: salários inferiores à realidade do mercado, ausência de um plano de benefícios, ambiente de trabalho inóspito, ausência de um plano de carreira e falta de um plano de desenvolvimento individual – PDI [3]. Uma das estratégias de retenção portanto, passa por uma remuneração adequada a realidade do mercado e às expectativas dos colaboradores.

O objetivo deste trabalho é explorar alguns modelos de *machine learning* que possibilitem prever salários com base nos dados do CAGED. Para isso o trabalho está dividido de seguinte forma: a seguir são apresentados os materiais e métodos utilizados para a seleção e tratamento dos dados bem como o treinamento e teste dos modelos. Em seguida, são apresentados os resultados dos treinamentos bem como um comparativo entre os modelos. Por fim, a última seção conclui.

2. Materiais e Métodos

Para este trabalho foram utilizados os dados de admissão do Cadastro Geral de Empregados e Desempregados – CAGED para o ano de 2022. O CAGED é um registro administrativo mantido pelo Ministério do Trabalho e Emprego – MTE e é utilizado pelo Programa de Seguro-Desemprego para conferência de dados de

vínculos trabalhistas, além de outros programas sociais. Os dados do CAGED são disponibilizados mensalmente pelo MTE e contém informações anonimizadas de movimentações – admissão e demissão – no mercado de trabalho formal ocorridas em todo território nacional. O registro da movimentação contém uma série de atributos, tais como: ano e mês da movimentação; localidade da movimentação (região, unidade federativa e município); atributos da empresa (seção e subclasse da Classificação Nacional de Atividade Econômica – CNAE e tamanho do estabelecimento); atributos do empregado e função desempenhada (código da ocupação de acordo com a Classificação Brasileira de Ocupações – CBO, grau de instrução, idade, raça/cor, sexo e tipo de deficiência); salário mensal declarado; e entre outros [2].

Os arquivos mensais contendo os dados foram extraídos do servidor do MTE via *File Transfer Protocol – FTP*, agrupados e armazenados em servidor SAS em arquivos de extensão *.sas7bdat* (tabelas SAS).

Os atributos originais utilizados da base foram: salário mensal (variável-alvo) e as variáveis explicativas (região, unidade federativa, município, seção CNAE, tamanho do estabelecimento, grau de instrução, raça/cor, sexo, idade, tipo de deficiência e ocupação CBO). Além destas, foram adicionados outros atributos à base, tais como: identificação *booleana* se o município é capital ou não, identificação *booleana* se o município está dentro do estado de São Paulo ou não, adição da microrregião do município com base na Divisão Territorial do Brasil – DTB do Instituto Brasileiro de Geografia e Estatística [5] e agrupamento da ocupação CBO em três níveis (família, subgrupo e grande grupo) [6]. Para selecionar os registros de admissão, utilizou-se apenas os dados que continham saldo de movimentação igual a um (1) e tipo de movimentação de códigos 10,20,25,35,70 e 97 que identificam respectivamente admissões por: primeiro emprego, reemprego, contrato de trabalho por prazo determinado, reintegração, transferência e tipo ignorado. Com esses tratamentos restaram na base 22,3 milhões de registros.

Foram necessários tratamentos adicionais nos dados para corrigir anomalias na média salarial de alguns meses e eliminar registros com atributos nulos ou não identificados. Na tabela abaixo são listadas as médias salariais para todos os meses de 2022 antes e depois da normalização. É possível perceber que para os meses de janeiro, fevereiro e dezembro a média salarial antes destoa dos outros meses. Tendo isso em vista, estabeleceu-se limites mínimos e máximos de salários

que seriam considerados. O limite mínimo estabelecido foi de R\$ 1.212,00 que é o valor do salário-mínimo vigente em 2022 e o máximo foi discricionariamente estabelecido em R\$ 150.000,00. É possível perceber que houve uma normalização da média salarial entre os meses após a adoção dos limites.

Tabela 1: Média salarial mensal, antes e depois da normalização dos dados, CAGED – 2022

Competência	Antes	Depois
202201	4.512,60	1.864,08
202202	133.398,51	1.810,82
202203	2.252,07	1.827,49
202204	2.286,21	1.809,71
202205	2.903,71	1.846,40
202206	2.223,05	1.867,77
202207	2.251,67	1.870,04
202208	2.885,15	1.899,84
202209	2.271,60	1.884,81
202210	2.173,42	1.896,96
202211	2.149,55	1.888,31
202212	25.492,89	1.874,18

Fonte: CAGED (2022), elaboração própria. [2]

Na tabela abaixo estão relacionados todos os filtros utilizados e seus respectivos impactos na exclusão de registros após a seleção das admissões. Foram consideradas apenas admissões com 44 horas semanais pois se refere a contratação padrão da Carteira de Trabalho e Previdência Social – CTPS. Com exceção dos filtros de horas contratuais e limites salariais, o restante dos filtros foi utilizado para excluir registros com atributos nulos ou não identificados. É possível perceber que os filtros com maiores impactos foram justamente os de horas contratuais e limites salariais respectivamente e que o restante dos filtros no máximo chegou a retirar 3,5% de registros da base o que evidencia que a maioria está com todos os atributos preenchidos. Ao final, restaram 16,2 milhões de registros.

Para o treinamento dos modelos, foi extraída uma amostra aleatória de 1% do total de 16,2 milhões, filtrando os registros com renda até oito salários-mínimos, que engloba 99,4% da população (TABELA 3). Essa amostra foi separada de forma estratificada com base na faixa de renda usando o pacote *StratifiedShuffleSplit* em 80% para treinamento e 20% para teste. A linguagem utilizada para treinamento foi *Python* versão 3.10.6, rodando em um computador pessoal com processador *AMD Ryzen 5-5500U* e 8GB de memória RAM. Os modelos testados foram: *DecisionTreeRegressor*, *GradientBoostingRegressor*, *RandomForestRegressor* e *GradientBoostingClassifier* da biblioteca *Scikit-Learn*. Todos os modelos foram treinados e aprimorados com o pacote *RandomizedSearchCV* para buscar, dentro de *ranges* pré-definidos, os melhores hiper parâmetros, a saber, aqueles que reduziram os erros na base de treinamento. Por fim, a métrica adotada para analisar os melhores hiper parâmetros e modelos foi o erro absoluto médio (*Mean Absolute Error – MAE*) para os regressores e a acurácia para os classificadores.

Tabela 2: Filtros aplicados à base de dados e respectivos impactos na exclusão de registros, CAGED – 2022

Filtro	Critério	Registros	% Excluídos
Registros		22.243.441	
Horas Contratuais	= 44	17.330.722	22,1%
Salário	>= 1.212 AND =< 150.000	20.528.517	7,7%
Tamanho Estabelecimento	=< 10	21.499.435	3,3%
Região	<> 9	22.212.541	0,1%
UF	<> 99	22.212.541	0,1%
CBO	<> 99.999	22.235.307	0,04%
Idade	IS NOT NULL	22.236.236	0,03%
Seção	<> Z	22.243.441	0,0%
Grau de Instrução	<> 99	22.243.441	0,0%
Raça/Cor	<> 9	22.243.441	0,0%
Sexo	<> 9	22.243.441	0,0%
Tipo Deficiência	<> 9	22.243.441	0,0%
Final		16.213.080	

Fonte: CAGED (2022), elaboração própria. [2]

GradientBoostingRegressor, *RandomForestRegressor* e *GradientBoostingClassifier* da biblioteca *Scikit-Learn*. Todos os modelos foram treinados e aprimorados com o pacote *RandomizedSearchCV* para buscar, dentro de *ranges* pré-definidos, os melhores hiper parâmetros, a saber, aqueles que reduziram os erros na base de treinamento. Por fim, a métrica adotada para analisar os melhores hiper parâmetros e modelos foi o erro absoluto médio (*Mean Absolute Error – MAE*) para os regressores e a acurácia para os classificadores.

Tabela 3: Distribuição de registros por faixa salarial, CAGED – 2022

Faixa de Renda	Quantidade	(%)
Mais de 1 a 3 SM	14.161.314	87,3%
1 SM	1.439.299	8,9%
Mais de 3 a 5 SM	398.692	2,5%
Mais de 5 a 8 SM	116.017	0,7%
Mais de 8 SM	97.758	0,6%
Total	16.213.080	100,0%

Fonte: CAGED (2022), elaboração própria. [2]

3. Resultados e Discussões

Como exploração inicial, treinou-se os modelos nos atributos ao nível de agregação máximo, tendo em vista as limitações de *hardware* – processamento e memória RAM – do equipamento utilizado. Os atributos utilizados para esta exploração inicial foram: grau de instrução, grande grupo CBO, idade, tamanho do estabelecimento, unidade federativa, seção da CNAE, sexo, raça/cor, região e tipo de deficiência. Os atributos agregados são: o grande grupo CBO, a seção CNAE e a região. Os resultados dos modelos treinados – *DecisionTreeRegressor*, *RandomForestRegressor* e *GradientBoostingRegressor* – seguem na tabela abaixo.

Tabela 4: Modelos treinados e respectivos resultados, treinamento com dados agregados

Modelo	Range Hiper par.	Melhores Hiper par.	MAE Train	MAE Test
DecisionTree	'max_depth':(1, 100) 'max_leaf_nodes':(1, 200)	'max_depth':7 'max_leaf_nodes':197	R\$ 342	R\$ 350
RandomForest	'n_estimators':(150, 160) 'max_depth':(1, 1000) 'max_leaf_nodes':(1, 1000)	'n_estimators':155 'max_depth':685 'max_leaf_nodes':560	R\$ 316	R\$ 342
GradientBoosting	'n_estimators':(49, 50) 'learning_rate':(0,01, 1,5) 'max_depth':(1, 120)	'n_estimators':49 'learning_rate':0.1165 'max_depth':10	R\$ 275	R\$ 335

Fonte: CAGED (2022), elaboração própria. [2]

Na tabela acima destaca-se duas informações. Em primeiro lugar, é possível perceber que os modelos tendem a generalização na medida em que, para todos, o erro absoluto médio na base de treinamento (*MAE – Treinamento*) é próximo ao erro absoluto médio na base de teste (*MAE – Teste*). Em segundo lugar, o modelo que apresentou melhor performance foi o *GradientBoostingRegressor* com erro médio na base de teste de R\$ 335,00. Os hiper parâmetros bem como o intervalo em que puderam variar durante o treinamento está descrito na coluna 'Range Hiper par' e os hiper parâmetros selecionados ao final dos treinamentos estão descritos na coluna 'Melhores Hiper par'.

Abaixo temos o resultado da *feature_importance_* do modelo *RandomForestRegressor* descrito acima. Percebe-se que os atributos que mais explicam o salário do empregado (mais de 50%) estão atrelados ao seu grau de instrução e ao trabalho (ocupação) que o empregado desempenha. Os atributos de localidade, como unidade federativa e região, explicam em conjunto cerca de 9% do valor dos salários.

Tabela 5: Resultado da *Feature_Importance_* do modelo *RandomForestRegressor*, treinamento dos modelos com dados agregados

Atributo	Feature Importance
Grau de Instrução	30,99%
Grande Grupo CBO	25,20%
Idade	12,95%
Tamanho Estabelecimento	8,52%
Unidade Federativa	8,30%
Seção CNAE	7,32%
Sexo	3,26%
Raça/Cor	2,22%
Região	1,15%
Tipo Deficiência	0,10%

Fonte: CAGED (2022), elaboração própria. [2]

Como segunda exploração, os modelos foram treinados em dados mais desagregados. O atributo unidade federativa foi substituído pelo município, a seção CNAE foi substituída pelo grupo CNAE e o grande grupo CBO substituído pela família CBO. Também se retirou o atributo tipo deficiência por ter baixa representatividade na *feature_importance_* e foram adicionados, mediante tratamento, os atributos: microrregião, substituindo a região; a identificação se o município é capital ou não (*Flag – Capital*); e se o município está dentro do estado de São Paulo ou não (*Flag – Estado São Paulo*). A hipótese levantada é de que uma vez agregados os atributos, os modelos treinados inicialmente perderam informações que pudessem fazê-los distinguir entre uma faixa salarial e outra. A hipótese estava certa e os resultados desta segunda rodada de treinamento seguem abaixo.

Tabela 6: Modelos treinados e respectivos resultados, treinamento com dados desagregados

Modelo	Range Hiper par.	Melhores Hiper par.	MAE Train	MAE Test
DecisionTree	'max_depth':(1, 100) 'max_leaf_nodes':(1, 200)	'max_depth':37 'max_leaf_nodes':88	R\$ 324	R\$ 332
RandomForest	'n_estimators':(150, 160) 'max_depth':(1, 1000) 'max_leaf_nodes':(1, 1000)	'n_estimators':153 'max_depth':193 'max_leaf_nodes':836	R\$ 272	R\$ 307
GradientBoosting	'n_estimators':(49, 50) 'learning_rate':(0,01, 1,5) 'max_depth':(1, 120)	'n_estimators':49 'learning_rate':0.1165 'max_depth':10	R\$ 224	R\$ 289

Fonte: CAGED (2022), elaboração própria. [2]

Destaca-se três informações. Em primeiro lugar, se observa, assim como na exploração inicial, que os modelos treinados nesta segunda rodada tendem também a generalização. Em segundo lugar, o que corrobora a hipótese descrita no parágrafo anterior é o fato de que o pior modelo treinado nesta segunda rodada é melhor que o melhor modelo treinado na exploração inicial. Em terceiro lugar, assim como na exploração inicial o modelo com a melhor performance foi o *GradientBoostingRegressor*. Abaixo, os resultados da *feature_importance_* do modelo *RandomForestRegressor* treinado nesta segunda rodada dão pistas de quais atributos mais contribuíram para a melhora do desempenho dos modelos. Em primeiro lugar, a CBO desagregada ao nível da família acrescentou mais informação aos modelos, esta alteração fez com que o atributo CBO saltasse de 25% para 41% da explicação sobre a predição dos salários, sendo o principal motivador da melhora. Em segundo lugar, o conjunto de localidade (município e microrregião) explicou melhor a predição de salários (11,5%) do que o conjunto anterior (uni-

dade federativa e região) que contribuiu com 9,5% da *feature_importance_*. Dos atributos adicionados à base o único que apresentou maior relevância para predição dos salários foi o atributo de microrregião. Os atributos que identificam se o município é capital e se está dentro do estado de São Paulo não apresentaram relevância, ficando com menos de 1% de importância cada.

Tabela 7: Resultado da *Feature_Importance_* do modelo *RandomForestRegressor*, treinamento com dados desagregados

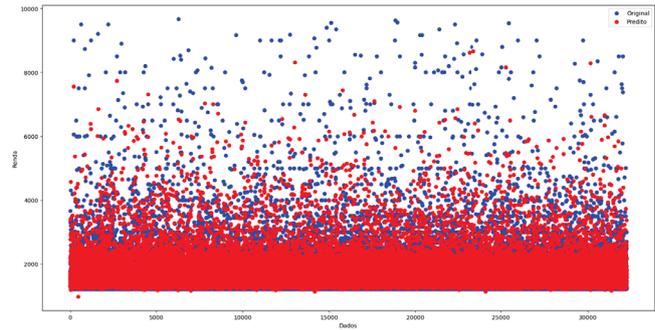
Atributo	Feature Importance
Família CBO	40,91%
Grau de Instrução	19,15%
Idade	9,01%
Grupo CNAE	7,96%
Município	7,74%
Tamanho Estabelecimento	6,96%
Microrregião	3,77%
Sexo	2,02%
Raça/Cor	1,42%
Flag - Estado de SP	0,68%
Flag - Capital	0,37%

Fonte: CAGED (2022), elaboração própria. [2]

Na imagem abaixo está representado o salário dos dados originais na base de teste (em azul) e os dados preditos (em vermelho) pelo modelo *GradientBoostingRegressor* com os dados desagregados. No eixo vertical tem-se o valor do salário e no horizontal apenas o index da base para se distribuir os dados. É possível perceber que o modelo prediz melhor salários mais baixos, até entorno dos R\$ 4.000 e para salários mais altos o modelo erra bastante. Isso se deve muito em conta a distribuição dos salários, a maior parte dos salários (96,2%) se concentra em até 3 salários-mínimos e a moda da base é R\$1.212, justamente o valor do salário-mínimo.

Tendo isso em vista, procurou-se usar um classificador *a priori* para determinar, se o registro, dado suas características, possuía um perfil de baixa renda ou de alta renda e então predizer o salário com base em um regressor treinado em baixas rendas ou um regressor treinado em altas rendas. O modelo geral dessa forma contaria com um classificador *a priori* e dois modelos de regressão *a posteriori* um para baixos salários e outro para altos salários.

Figura 1: Salários originais e preditos do modelo *GradientBoostingRegressor*



Fonte: CAGED (2022), elaboração própria. [2]

Para treinar o classificador era preciso estipular um valor que dividisse a base em salários baixos e salários altos.

De forma inicial, estabeleceu-se o valor em R\$ 4.000 que é, pelo gráfico acima, a renda a partir da qual os erros começam a serem mais evidentes. O classificador teve uma boa acurácia na base de teste para classificar baixas rendas, com 99% de acerto, porém um péssimo desempenho para acertar quando os registros estavam em altas rendas, com apenas 27% de acurácia. O modelo geral, com os preditores de baixos e altos salários apresentou *MAE* na base de teste de R\$ 289,00 que é próximo ao erro médio apresentado pelo modelo *GradientBoostingRegressor* treinado com os dados desagregados. Isso se deve muito ao bom desempenho nas baixas rendas que compõem a maior parte da base (acurácia de 99% e *MAE* de R\$ 217,00).

Em uma segunda rodada, na tentativa de melhorar a acurácia da classificação para altas rendas, utilizou-se o recorte em R\$ 1.800 para classificação que é o valor que divide a base entre 75% (os menores salários) e os 25% restantes (maiores salários). A classificação melhorou, saltando dos iniciais 27% para 65% de acurácia para rendas altas, porém, ao custo de uma redução da acurácia na classificação das rendas baixas de 99% para 90%. Esta redução da acurácia para as baixas rendas foi crucial para aumentar o *MAE* do modelo geral que apresentou valor médio de R\$ 301, sendo este o segundo melhor modelo treinado, porém, ainda perdendo em desempenho para o *GradientBoostingRegressor* treinado com os dados desagregados.

4. Conclusão

Treinar modelos de *machine learning* é um processo de exploração. Sobre os dados do CAGED, tendo em

vista os recursos de *hardware* disponíveis, foi possível explorar modelos mais simples com dados não tão desagregados – como a ocupação CBO – que exigiram baixa demanda computacional.

Nos modelos explorados o que fica patente é que a distribuição dos salários majoritariamente concentrados em faixas mais baixas fez tender para baixo as predições e portanto, os modelos não desempenharam bem na predição de altas rendas, dada a pouca representatividade destas na base. Também se destaca que independente do tratamento e dos atributos utilizados da base – desagregados ou mais agregados – os que mais explicaram a diferença salarial entre os empregados foram a ocupação, o grau de instrução e a idade. Dentre os modelos analisados, o que obteve o menor erro médio foi o *GradientBoostingRegressor*.

Para futuros trabalhos sugere-se aprimorar o classificador, adotando para isso outros métodos como redes neurais, utilizar técnicas de *oversampling*, se necessário e treinar modelos quantílicos. A adição de outros atributos pode ajudar na melhoria da predição. Com a Lei de Acesso à Informação, é possível solicitar a adição de novos atributos à base que possam estar de posse do Ministério do Trabalho e Emprego e que não são divulgados na base atual. Dados estes que podem enriquecer a base sem, contudo, comprometer o princípio da anonimização dos dados.

Referências

- [1] AL-SURAIHI, Walid A *et al.* **Employee Turnover: Causes, Importance and Retention Strategies.** *European Journal of Business Management and Research*, 2021. Disponível em: <https://www.ejbmr.org/index.php/ejbmr/article/view/893>. Acesso em: jun, 2023.
- [2] CAGED. **Cadastro Geral de Empregados e Desempregados.** 2022. Disponível em: ftp.mtps.gov.br/pdet/microdados/NOVO_CAGED/2022/. Acesso em: jan, 2023.
- [3] Fortes Tecnologia. **Turnover: Por que pode ser tão prejudicial para as empresas?** 2020. Disponível em: <https://blog.fortestecnologia.com.br/gestao-pessoas/turnover-pode-ser-prejudicial/>. Acesso em: jun, 2023.
- [4] GRANVILLE. Lisandro Z. **Parceria da Revolução Digital.** *Revista da Sociedade Brasileira de Computação*. n° 39. Porto Alegre. 2019. Disponível em: <https://www.sbc.org.br/publicacoes-mainmenu/computacao-brasil>. Acesso em: jun, 2023.
- [5] IBGE. Instituto Brasileiro de Geografia e Estatística. **Divisão Territorial do Brasil.** 2023. Disponível em: [ofp.igge.gov.br/organizacao_do_territorio/estrutura_territorial/divisao_territorial/2022](https://ftp.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/divisao_territorial/2022). Acesso em: abr, 2023.
- [6] MTE. Ministério do Trabalho e Emprego. **Estrutura CBO: Classificação Brasileira de Ocupações.** 2023. Disponível em: <https://cbo.mte.gov.br/cbsite/pages/downloads.jsf>. Acesso em: abr, 2023.
- [7] Robert Half. **Turnover em alta: Principais motivos e como evitar.** 2022. Disponível em: <https://www.roberthalf.com.br/blog/tendencias/turnover-em-alta-principais-motivos-e-como-evitar>. Acesso em: jun, 2023.
- [8] SINPROFAZ. Sindicato Nacional dos Procuradores da Fazenda Nacional. **Quanto Custa o Brasil: SONEGÔMETRO.** 2019. Disponível em: <http://www.quantocustaobrasil.com.br>. Acesso em: jun, 2023.