

TELETRABALHO NO CONTEXTO BRASILEIRO: FATORES SOCIAIS E ECONÔMICOS SOB A ABORDAGEM DE MACHINE LEARNING

Laryssa de Andrade Mairinque – laryssamairinque@unifei.edu.br

Robson Bruno Dutra Pereira – robsondutra@ufs.br

Josiane Palma Lima – jplima@unifei.edu.br

Palavras-chave: Teletrabalho, Machine Learning, Brasil, Variáveis Socioeconômicas.

1. INTRODUÇÃO

No passado, a natureza das ferramentas de trabalho ditava o local físico onde as tarefas precisavam ser realizadas. No entanto, na era digital atual, muitas dessas ferramentas tornaram-se portáteis, permitindo que determinadas categorias de trabalho sejam executadas virtualmente em qualquer lugar (Hopkins & McKay, 2019). O teletrabalho refere-se à realização de atividades profissionais fora de escritórios tradicionais ou locais fixos de trabalho. Do ponto de vista ambiental, o potencial de substituir viagens motorizadas por “viagens virtuais” é particularmente atraente (Martens & Korver, 2000).

Nesse contexto, o teletrabalho surge como um substituto viável para certos deslocamentos diários, permitindo que as pessoas realizem atividades profissionais e pessoais remotamente. Assim, ele tem sido cada vez mais visto como uma ferramenta de política pública com potencial para aliviar a congestão viária, reduzir custos relacionados ao transporte e mitigar as emissões de gases de efeito estufa (Mouratidis & Peters, 2022). Desta forma, é importante entender quais as variáveis influenciam na realização do teletrabalho no contexto nacional.

Diante desse cenário, este estudo tem como objetivo modelar o regime de trabalho adotado pelos trabalhadores brasileiros, trabalho presencial ou remoto, a partir de variáveis socioeconômicas, utilizando uma abordagem de *machine learning*. Os dados foram obtidos da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Essa pesquisa é uma das principais fontes de informação sobre as condições

socioeconômicas da população brasileira, fornecendo dados atualizados sobre mercado de trabalho, educação e rendimento. De caráter amostral e probabilístico, a PNAD Contínua é realizada de forma contínua em todo o território nacional, abrangendo áreas urbanas e rurais. A amostra, composta por cerca de 210 mil domicílios em aproximadamente 3.500 municípios, garante ampla representatividade e robustez estatística aos resultados obtidos.

2. DESENVOLVIMENTO

A adoção do teletrabalho pode ser influenciada por uma variedade de fatores. De modo geral, a literatura indica que o perfil típico do teletrabalhador é predominantemente composto por homens (López Soler, Christidis & Vassallo, 2021; Ravalet & Rérat, 2019), com altos níveis de renda (Sweet & Scott, 2022), maior nível de escolaridade (Ravalet & Rérat, 2019; Sweet & Scott, 2022) e, geralmente, pertencente à faixa etária de meia-idade (Ravalet & Rérat, 2019). Esses fatores sugerem que o acesso ao teletrabalho tende a ser mais comum entre indivíduos com maior capital social e econômico, o que pode reforçar desigualdades no contexto do trabalho remoto.

Alguns estudos utilizaram métodos de *machine learning*, particularmente algoritmos baseados em árvores, para prever a adoção ou a preferência pelo trabalho remoto. Esses modelos demonstraram alta precisão na identificação de padrões e preferências relacionadas ao trabalho remoto (Setiawan & Alamsari, 2022). No entanto, a maioria desses estudos não explorou sistematicamente o impacto das variáveis sociodemográficas em suas previsões. Apenas dois estudos incluíram tais variáveis: um relatou idade, gênero e função profissional, mas não avaliou seu poder preditivo (Chen et al., 2023), enquanto o outro mencionou apenas a variável relacionada à função profissional (Rehan et al., 2023). Assim, a influência explícita das características sociodemográficas, ou de outros fatores relacionados, permanece amplamente inexplorada nesse contexto.

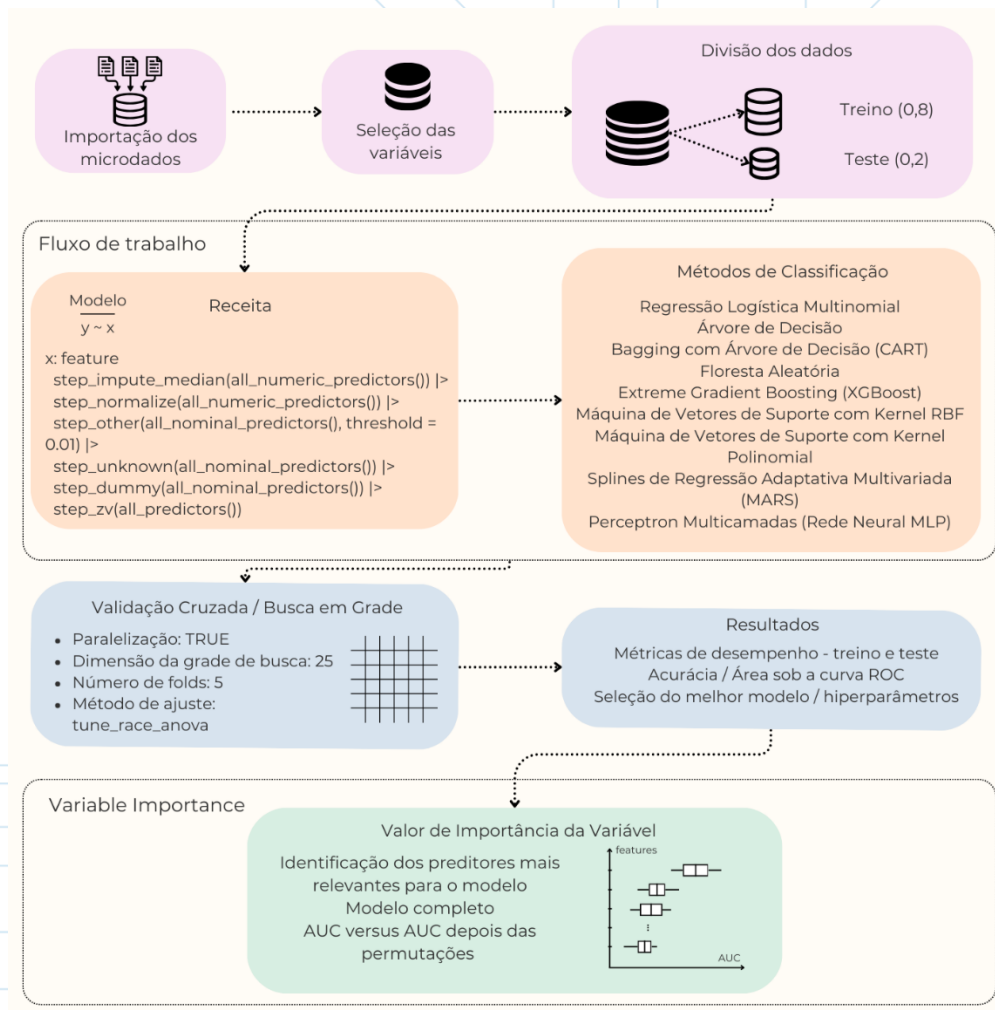
A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), é uma das principais fontes de informação sobre as condições socioeconômicas da população

brasileira. No primeiro trimestre de 2025, a pesquisa estimou uma população de 177,17 milhões de pessoas com 14 anos ou mais, das quais 110,64 milhões (62,2%) integravam a força de trabalho e 66,98 milhões (37,8%) estavam fora dela. Nesse período, havia aproximadamente 103,8 milhões de pessoas ocupadas (58,6% da população com 14 anos ou mais) e 6,82 milhões de desocupadas (3,8%), resultando em uma taxa de desocupação de 6,2% e uma taxa de participação na força de trabalho de 62,2% (IBGE, 2025).

A Figura 1 apresenta as etapas do processo de aprendizado supervisionado e de análise de importância das variáveis utilizadas neste estudo. Inicialmente, os microdados da PNAD Contínua de 2025 foram carregados e organizados para análise. Após a seleção das variáveis, os dados foram divididos em conjuntos de treinamento (80%) e teste (20%). A variável alvo (y) representa o regime de trabalho realizado (remoto ou presencial), enquanto as demais variáveis foram utilizadas como preditoras (x1 a xn).

Foram testados diversos métodos de classificação, com os hiperparâmetros ajustados via validação cruzada (5 folds) com grid search. O desempenho dos modelos foi avaliado pela acurácia e pela Área sob a curva ROC no treinamento e teste. O modelo final foi selecionado considerando o menor erro na validação e no conjunto de teste. A importância das variáveis foi avaliada embaralhando-se cada preditor e medindo o impacto no desempenho, visualizado por boxplots.

FIGURA 1 Etapas de trabalho



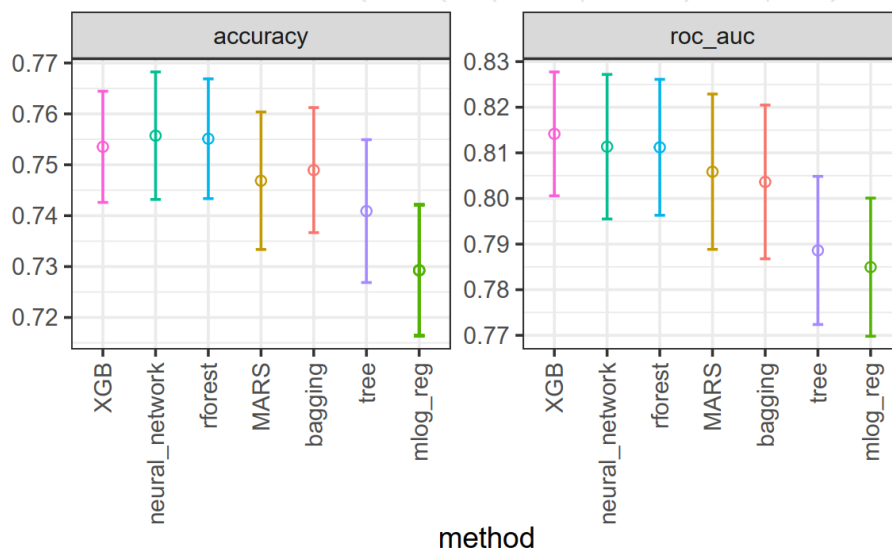
Fonte: Autoria própria

3. RESULTADOS E DISCUSSÃO

A Figura 2 apresenta os intervalos de confiança das métricas de desempenho accuracy e ROC AUC para os diferentes métodos de classificação avaliados por validação cruzada. Cada ponto representa a média das métricas nos cinco folds, enquanto as linhas verticais indicam os intervalos de confiança. Observa-se que os métodos XGB, neural_network e rforest apresentaram os melhores resultados tanto em accuracy quanto em ROC AUC, com médias mais altas e intervalos de confiança

menores, indicando maior estabilidade. Já métodos como tree e mlog_reg apresentaram desempenho inferior e maior variabilidade.

FIGURA 2 Intervalo de confiança da validação cruzada de 5 folds para os métodos classificação

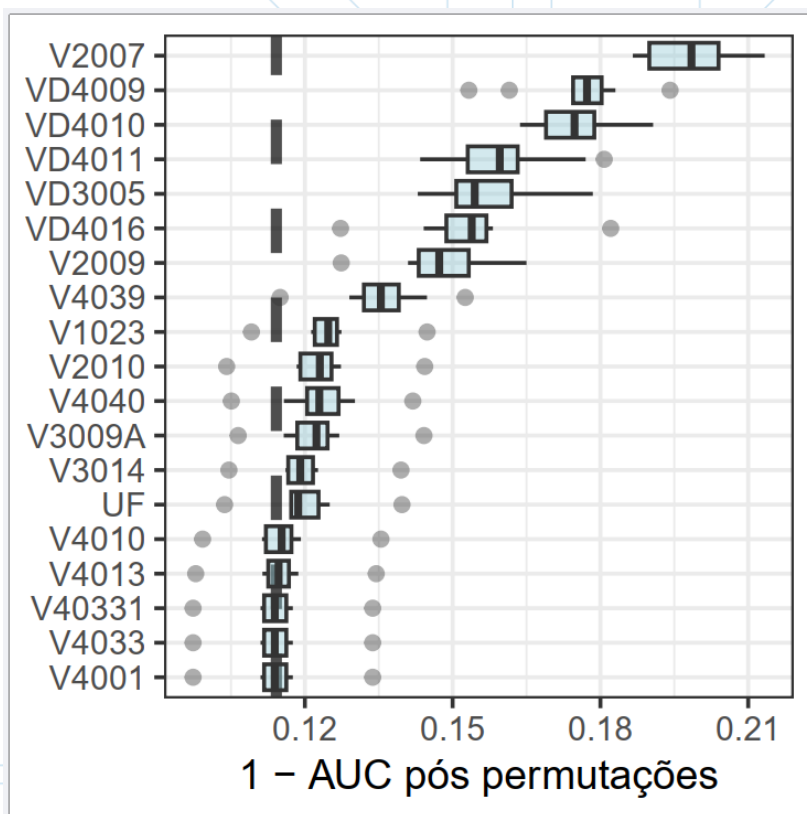


Fonte: Autoria própria

O modelo XGB com os hiperparâmetros otimizados apresentou a melhor performance, com um valor de acurácia igual a 0,769 e um valor de ROC AUC igual a 0,831. Posteriormente foram avaliadas a importância das variáveis preditivas no modelo conforme apresentado na Figura 3. A variável de maior importância para o modelo foi a V2007 que é referente ao sexo dos indivíduos entrevistados. Em seguida destacam-se as variáveis VD4009 e VD4010, que são a posição na ocupação e categoria do emprego do trabalho, e os grupamentos de atividade principal do empreendimento do trabalho principal. Ambas variáveis relacionadas à natureza do trabalho realizado. Variáveis de outras naturezas como Anos de estudo (VD3005), Rendimento mensal (VD4016), Idade (V2009) também foram relevantes para a variável dependente estudada, que é o modo de realização de trabalho no Brasil.



FIGURA 3 Importância das variáveis preditivas no modelo XBG



Fonte: Autoria própria

4. CONSIDERAÇÕES FINAIS

O presente estudo evidencia que a adoção do teletrabalho no Brasil está fortemente associada às características sociodemográficas e ocupacionais dos indivíduos. Ao aplicar os métodos de *machine learning*, foi possível identificar com maior precisão as variáveis mais relevantes para prever o regime de trabalho adotado. Entre elas, destacam-se o sexo do trabalhador, a posição e categoria na ocupação, o grupamento de atividade do empreendimento, os anos de estudo, a idade e o rendimento mensal. Esses resultados indicam que tanto fatores sociodemográficos quanto características do trabalho exercido têm papel determinante na adoção do teletrabalho no contexto brasileiro.

Os resultados reforçam a necessidade de políticas públicas e organizacionais que promovam maior equidade no acesso ao trabalho remoto e apontam para a utilidade de técnicas de *machine learning* na análise de grandes bases de dados nacionais. Pesquisas futuras podem incorporar fatores qualitativos e contextuais para compreender melhor a dinâmica do teletrabalho no Brasil.

AGRADECIMENTOS

Os autores agradecem à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro concedido aos projetos que subsidiaram o desenvolvimento deste trabalho.

REFERÊNCIAS

Hopkins, J. L., & McKay, J. (2019). Investigating 'anywhere working' as a mechanism for alleviating traffic congestion in smart cities. *Technological Forecasting and Social Change*, 142, 258–272.

IBGE – Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional por Amostra de Domicílios Contínua: quadro sintético 1º trimestre de 2025 (jan-fev-mar)*. Coordenação de Trabalho e Rendimento, 2025.

López Soler, J. R., Christidis, P., & Vassallo, J. M. (2021). Teleworking and online shopping: socio-economic factors affecting their impact on transport demand. *Sustainability*, 13(13), 7211.

Martens, M., & Korver, W. (2000). Forecasting and assessing the mobility effects of teleservices: scenario approach. *Transportation Research Record*, 1706(1), 118–125.

Mouratidis, K., & Peters, S. (2022). Covid-19 impact on teleactivities: Role of built environment and implications for mobility. *Transportation Research Part A: Policy and Practice*, 158, 251–270.



IEPG SUMMIT'25

Pensando o futuro com inteligência
artificial e consciência social

Ravalet, E., & Rérat, P. (2019). Teleworking: decreasing mobility or increasing tolerance of commuting distances? *Built Environment*, 45(4), 582–602.

Setiawan, J., & Alamsari, R. G. (2022, dec 8). Prediction of Work From Home Post COVID-19 using Classification Model. In *2022 Seventh International Conference on Informatics and Computing (ICIC)* (pp. 1–6).

Sweet, M., & Scott, D. M. (2022). Insights into the future of telework in canada: Modeling the trajectory of telework across a pandemic. *Sustainable Cities and Society*, 87, 104175.