



ANÁLISE PREDITIVA DA DEMANDA REGIONAL POR VAGAS DE TI COM APRENDIZADO DE MÁQUINA

Danilo Batista Lima¹, Isabelle Yasmin de Araujo Moreira², Carlos Alexandre Silva³

RESUMO

Este estudo desenvolve e compara modelos de aprendizado de máquina para a previsão da demanda regional por vagas no setor de Tecnologia da Informação (TI) no Brasil. A partir de um conjunto de dados com mais de 66 mil anúncios de emprego, extraídos do LinkedIn via *web scraping* e estruturados como séries temporais diárias, a análise abrange as cinco macrorregiões brasileiras e uma categoria específica para vagas remotas. Foram avaliados comparativamente três algoritmos de aprendizado de máquina para previsão, sendo eles *Support Vector Regression* (SVR), redes neurais *feedforward* do tipo *Multi-Layer Perceptron* (MLP) e redes neurais *Long Short-Term Memory* (LSTM). O desempenho dos modelos, treinados individualmente por região, foi aferido por meio das métricas MAE (*Mean Absolute Error*) e MDA (*Mean Directional Accuracy*). Os resultados demonstram que o modelo *feedforward* obteve a melhor performance, apesar do modelo SVR ter obtido resultados parecidos. De acordo com os dados, as vagas remotas, somadas às das regiões Sudeste e Sul, representam mais de 90% do total, evidenciando a menor oferta nas demais regiões. Dessa forma, este trabalho demonstra a viabilidade da metodologia para a captura de padrões de demanda regionais, oferecendo um panorama quantitativo robusto do comportamento do setor de TI brasileiro por região e modalidade de trabalho.

Palavras-chave: Aprendizado de Máquina. Mercado de Trabalho de TI. Previsão de Séries Temporais.

1 INTRODUÇÃO

O setor de Tecnologia da Informação (TI) no Brasil, o maior da América Latina (ABES, 2024), exibe uma acentuada desigualdade na distribuição de oportunidades, com forte concentração histórica na região Sudeste. A recente expansão do trabalho remoto, no entanto, tem reconfigurado esse cenário, criando um ambiente propício para a interiorização de vagas e a redução das disparidades regionais. Nesse contexto, a previsão da demanda por profissionais de TI por região e modalidade de trabalho torna-se uma ferramenta estratégica para a análise aprofundada deste mercado.

Embora a literatura internacional apresente estudos sobre o uso de aprendizado de máquina para prever tendências no mercado de trabalho (SENTHURVELAUTHAM,

¹ Bacharelado em Sistemas de Informação, Campus Sabará, IFMG

² Bacharelado em Sistemas de Informação, Campus Sabará, IFMG

³ Doutor em Ciência da Computação e Matemática Computacional, Campus Sabará, IFMG



SENANAYAKE, 2023; GAJEWSKI et al., 2023), são escassos os trabalhos que exploram o tema no contexto regional brasileiro. As pesquisas existentes sobre o mercado nacional, em geral, não detalham diferenças regionais ou não utilizam modelos preditivos voltados para a abertura de vagas no setor de tecnologia (TESSARIN, MORCEIRO, 2022; BRITTO et al., 2023).

Este estudo propõe uma análise preditiva da oferta de vagas de TI por região, utilizando o LinkedIn como fonte de dados. Foram comparados três algoritmos de aprendizado de máquina: *Support Vector Regression* (SVR), redes neurais *feedforward* do tipo *Multi-Layer Perceptron* (MLP) e *Long Short-Term Memory* (LSTM), avaliando o desempenho de cada um por meio das métricas MAE (*Mean Absolute Error*) e MDA (*Mean Directional Accuracy*) para identificar o modelo mais eficaz na previsão da abertura de vagas.

2 DESENVOLVIMENTO

2.1 Metodologia

A metodologia foi executada em três fases: coleta, pré-processamento e treinamento dos modelos. Primeiramente, realizou-se a coleta de dados por meio de *web scraping* no LinkedIn, onde um algoritmo em Python extraiu diariamente, entre 1 de março e 1 de junho de 2025, anúncios de vagas de TI. Vale salientar que foram coletados apenas dados públicos e não pessoais, conforme a Lei Geral de Proteção de Dados (CARDOSO, 2021). Após limpeza e remoção de vagas irrelevantes, a base de dados final continha 66.863 registros. Estes dados, por sua vez, foram classificados em 6 categorias de acordo com a região e modalidade: Norte (0.9% das vagas da base), Nordeste (5.2%), Centro-Oeste (3.7%), Sudeste (32.3%), Sul (20.3%) e Remoto (37.6%).

Então, os dados foram pré-processados e agregados em seis séries temporais diárias (cinco macrorregiões e a modalidade remota), com tratamento de *outliers* via método interquartil e normalização Min-Max. Por fim, foram treinados e avaliados três algoritmos de aprendizado de máquina (MLP, LSTM e SVR), com instâncias específicas para cada série temporal. Utilizando o dia da semana e *lags* como *features*, os modelos tiveram seus hiperparâmetros otimizados via *Grid Search* e seu desempenho foi aferido com as métricas MAE e MDA, permitindo avaliar tanto a precisão absoluta das previsões quanto a capacidade



de capturar corretamente a direção das variações no número de vagas. O conjunto de teste utilizado correspondia a 20% dos dados.

2.1 Resultados

Os resultados da avaliação são apresentados na Tabela 1. Todas as métricas foram calculadas com base no desempenho obtido no conjunto de teste.

Tabela 1 - Métricas obtidas após treinamento dos algoritmos para o conjunto de dados de teste

Métrica	Feedforward	LSTM	SVR
MDA (%)	80.7%	62.2%	74.0%
MAE	30.4	42.5	29.2

Fonte: Elaborado pelos autores (2025).

As métricas indicam que o modelo *feedforward* alcançou o melhor desempenho médio no conjunto de teste, com um MAE de 30.4 e um MDA de 80.7%. Embora o SVR tenha obtido um MAE ligeiramente menor (29.2), sua capacidade de prever a direção da tendência (MDA de 74.0%) foi inferior à do *feedforward*. O modelo LSTM apresentou o desempenho inferior em ambas as métricas (MAE de 42.5 e MDA de 62.2%).

Os gráficos do apêndice A também confirmam a performance do modelo *feedforward*. Percebe-se que captura de forma mais consistente a dinâmica de picos e vales das séries, especialmente nas categorias com maior volume de dados (Remoto, Sudeste e Sul), o que reflete seu elevado MDA. O SVR demonstrou um desempenho próximo, justificado pelo uso do *kernel* RBF, que lhe confere capacidade de modelar padrões não lineares. Em contrapartida, o LSTM tendeu a suavizar as previsões, falhando em replicar a volatilidade dos dados, possivelmente por sua arquitetura privilegiar padrões de longo prazo em detrimento de variações de curto prazo, fator que afetou suas métricas de desempenho.

3 CONCLUSÃO

Este estudo demonstrou a viabilidade da aplicação de modelos de aprendizado de máquina para prever a demanda regional por vagas de TI no Brasil. A metodologia de coleta de dados do LinkedIn e o treinamento de modelos específicos por região se mostraram eficazes para capturar padrões complexos do mercado.

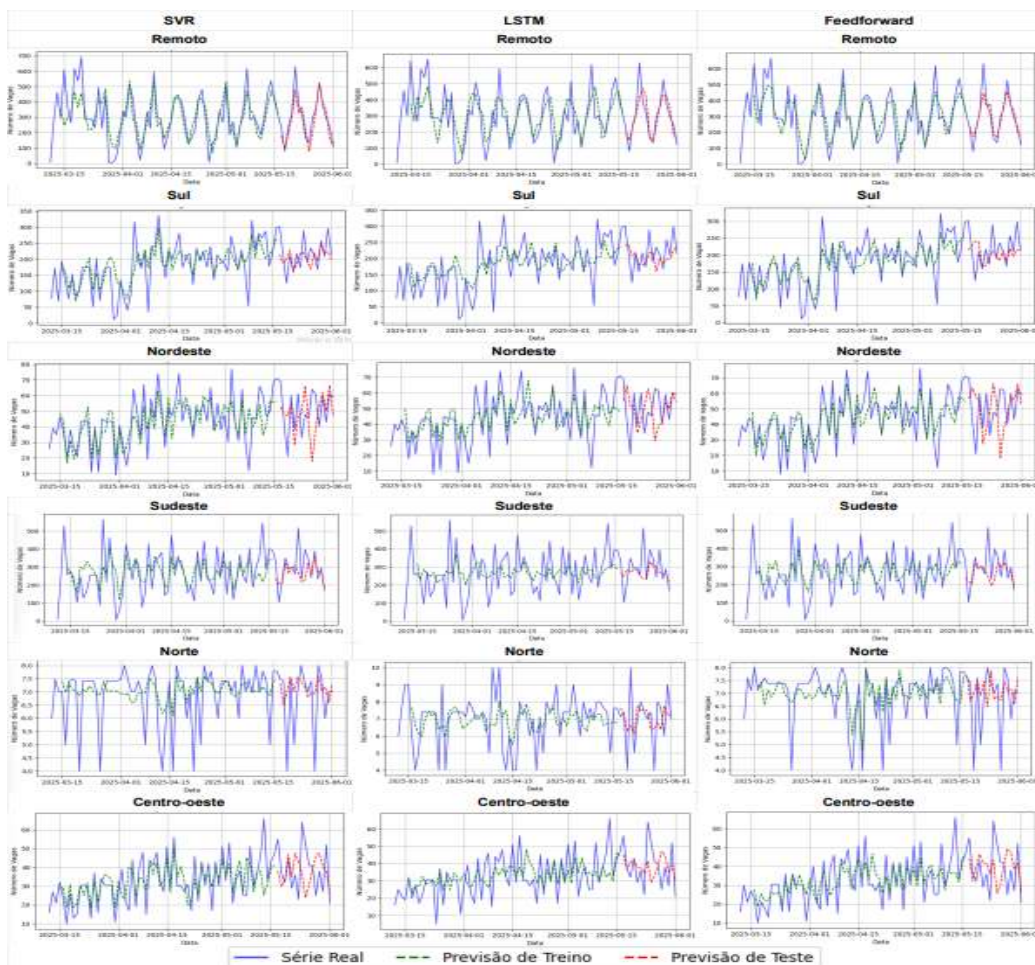
Com base nos resultados, observou-se que o mercado de TI no Brasil apresenta forte concentração de vagas, com a modalidade remota, a região Sudeste e a Sul respondendo, em



conjunto, por mais de 90% das oportunidades, enquanto as demais regiões possuem participação significativamente menor. Em termos de desempenho preditivo, a rede neural *feedforward* apresentou-se como o modelo mais equilibrado e robusto, obtendo o melhor balanço entre a minimização do erro (MAE) e a acurácia direcional (MDA). O SVR surgiu como uma alternativa competitiva, enquanto a rede LSTM se mostrou menos adequada para séries temporais com a alta volatilidade observada.

Este trabalho contribui para a literatura ao oferecer uma análise preditiva da demanda regional no mercado de TI brasileiro. Como sugestões para trabalhos futuros, recomenda-se a ampliação do horizonte temporal da base de dados e a incorporação de variáveis exógenas, como indicadores macroeconômicos, para aprimorar a capacidade preditiva dos modelos. Os autores agradecem o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fomento que tornou viável esta pesquisa.

APÊNDICE A – Gráficos das previsões realizadas pelos modelos





REFERÊNCIAS

ABES: Associação Brasileira das Empresas de Software. **Mercado brasileiro de software: panorama e tendências, 2024**. Disponível em: <https://abes.org.br/dadosdo-setor/>. Acesso em 18 set. 2024.

CARDOSO, O. V. O web scraping viola a proteção de dados pessoais? **JusBrasil**, 2021. Disponível em: <https://www.jusbrasil.com.br/artigos/o-web-scrapingviola-a-protecao-de-dados-pessoais/1152362639>. Acesso em 22 dez. 2024.

BRITTO, J.; URRACA-RUIZ, A.; FERRAZ, J.; TORRACCA, J.; LACERDA, H. El impacto de la digitalización sobre empleo y las habilidades por estadios de adopción en Brasil y Argentina. **Revista Brasileira de Inovação**, 2023.

GAJEWSKI, Przemysław; ČULE, Bojan; RANKOVIC, Nikola. Unveiling the Power of ARIMA, Support Vector and Random Forest Regressors for the Future of the Dutch Employment Market. **Journal of Theoretical and Applied Electronic Commerce Research**, v. 18, n. 3, p. 1365-1403, 2023. DOI: 10.3390/jtaer18030069.

SENTHURVELAUTHAM, Sharanaja; SENANAYAKE, Nipuna. A machine learning based job forecasting and trend analysis system to predict future job markets using historical data. **2023 IEEE 8TH INTERNATIONAL CONFERENCE FOR CONVERGENCE IN TECHNOLOGY (I2CT)**, 8., 2023. p. 1-7. DOI: 10.1109/I2CT57861.2023.10126233.

TESSARIN, Milene; MORCEIRO, Paulo César. Labour market transformations in the era of new technologies: an analysis by regions, gender and industries in Brazil. **Economic Research Southern Africa**, 2022. DOI: 10.54223/uniwitwatersrand10539-33455