

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

Modelagem e Previsão de Consumo Energético Urbano: Aplicação de Técnicas de Aprendizagem de Máquina na Análise de Dados em Ambientes Urbanos

VINICIUS S. SANTOS¹, MURILO V. SILVA²

¹Graduando em Engenharia da Computação, Bolsista FAPESP, IFSP, Campus Birigui, vinicius.santos@ifsp.edu.br

²Doutor em Ciências da Computação, Docente na Área da Informática, IFSP, Campus Birigui, murilo.varges@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação.

RESUMO: Este estudo investigou a aplicação de técnicas de aprendizado de máquina, particularmente a Floresta Aleatória (*Random Forest*), na previsão do consumo energético urbano na cidade de Tetuão, Marrocos. A análise comparativa entre diversos modelos preditivos, incluindo Regressão Linear, K-Nearest Neighbors e Gradient Boosting, demonstrou a superioridade da Floresta Aleatória em termos de precisão e robustez. A inclusão de variáveis climáticas e temporais no modelo foi crucial para a obtenção de previsões mais precisas, validando a hipótese central do estudo. Os resultados reforçam a eficácia da Floresta Aleatória como uma ferramenta valiosa para a gestão energética e o planejamento de infraestrutura em ambientes urbanos. Este trabalho contribui significativamente para a literatura, oferecendo uma abordagem prática para a previsão de consumo energético em cidades inteligentes.

PALAVRAS-CHAVE: Previsão de consumo energético; Floresta Aleatória; aprendizado de máquina; dados urbanos; planejamento urbano; cidades inteligentes.

Modeling and Forecasting of Urban Energy Consumption: Application of Machine Learning Techniques in Data Analysis in Urban Environments

ABSTRACT: This study investigated the application of machine learning techniques, particularly Random Forest, in predicting urban energy consumption in the city of Tetouan, Morocco. The comparative analysis between several predictive models, including Linear Regression, K-Nearest Neighbors, and Gradient Boosting, demonstrated the superiority of Random Forest in terms of accuracy and robustness. The inclusion of climate and temporal variables in the model was crucial for obtaining more accurate predictions, validating the central hypothesis of the study. The results reinforce the effectiveness of Random Forest as a valuable tool for energy management and infrastructure planning in urban environments. This work contributes significantly to the literature by offering a practical approach for predicting energy consumption in smart cities.

KEYWORDS: Energy consumption forecasting; Random Forest; machine learning; urban data; urban planning; smart cities.

INTRODUÇÃO

As rápidas transformações urbanas, impulsionadas pelo crescimento populacional e inovações, destacam a importância da previsão do consumo de energia para o desenvolvimento sustentável das cidades (EPE, 2023). Técnicas de *machine learning*, como a Floresta Aleatória (*Random Forest*), têm se mostrado particularmente eficazes na compreensão de padrões complexos em grandes volumes de dados (BARTOLO, 2018; GODINHO, 2020). Este projeto visa desenvolver um modelo preditivo utilizando a Floresta Aleatória para prever o consumo de energia na cidade de Tetuão, Marrocos, com o objetivo específico de identificar padrões sazonais e comportamentais que possam otimizar a gestão energética urbana.

Estudos recentes (SHAPI; RAMLI; AWALIN, 2021; SANTOS; CHAUCOSKI, 2020) reforçam o papel crescente do *machine learning* na previsão de consumo energético em ambientes urbanos, sendo essencial para a gestão eficiente de recursos em cidades inteligentes. Ademais, técnicas como Floresta Aleatória e Gradient Boosting revelam padrões ocultos nos dados, oferecendo insights valiosos para a formulação de políticas públicas e decisões estratégicas em planejamento urbano (NALLATHAMBI; RAMASAMY, 2017; AMARAL, 2019). A hipótese central deste estudo é que a integração de variáveis climáticas e temporais ao modelo de Floresta Aleatória não apenas resultará em previsões mais precisas em comparação com métodos tradicionais, mas também permitirá uma adaptação dinâmica a diferentes cenários climáticos e padrões de uso, o que é crucial para a tomada de decisões estratégicas em planejamento urbano.

A previsão precisa do consumo de energia não apenas aprimora a eficiência energética, mas também desempenha um papel fundamental na alocação de recursos e no planejamento urbano sustentável. Essa abordagem proporciona uma base sólida para o desenvolvimento das cidades em um mundo cada vez mais dependente de energia e recursos (GODINHO, 2020; ZOGAAN, 2022).

MATERIAIS E MÉTODOS

Para atingir o objetivo de prever o consumo energético na cidade de Tetuão, Marrocos, com alta precisão, foi necessário desenvolver um modelo utilizando técnicas de aprendizado de máquina. A base de dados utilizada neste estudo foi obtida do trabalho de (SALAM; HIBAOUI, 2018), que fornece dados detalhados de consumo energético coletados pela Amendis através do sistema SCADA (Supervisory Control and Data Acquisition). Esses dados foram fundamentais para a construção e validação do modelo proposto.

Importação e Pré-Processamento dos Dados

A primeira etapa consistiu na importação dos dados brutos, seguida de um pré-processamento. Esta fase incluiu a detecção e remoção de *outliers*, que poderiam distorcer os resultados preditivos, e a filtragem de dados para eliminar ruídos. Valores ausentes foram tratados para evitar lacunas que poderiam comprometer a integridade das análises subsequentes. Adicionalmente, a coluna de data e hora (DateTime) foi ajustada para garantir a correta alocação temporal dos eventos, o que é crucial para análises temporais e sazonais.

Análise Exploratória de Dados

Com os dados pré-processados, foi realizada uma Análise Exploratória de Dados (AED) para compreender a distribuição das variáveis e os padrões subjacentes. A AED envolveu a aplicação de métodos estatísticos descritivos, análise de séries temporais, e avaliação de correlações entre variáveis. Boxplots foram gerados para identificar a presença de *outliers*, e a análise de padrões temporais revelou como as variáveis comportam-se ao longo do tempo. Essa etapa foi essencial para informar as decisões subsequentes na modelagem.

Modelagem

O modelo preditivo principal foi desenvolvido utilizando a técnica de Floresta Aleatória, escolhida devido à sua comprovada eficácia em lidar com grandes volumes de dados e variáveis complexas (BREIMAN, 2001). A Figura 1 ilustra o processo de construção do modelo de Floresta Aleatória, que consiste em múltiplas árvores de decisão que votam para determinar a classe final da previsão. Este método é particularmente eficaz para problemas de regressão como o abordado neste estudo, pois combina as previsões de várias árvores de decisão para reduzir a variabilidade e melhorar a precisão.

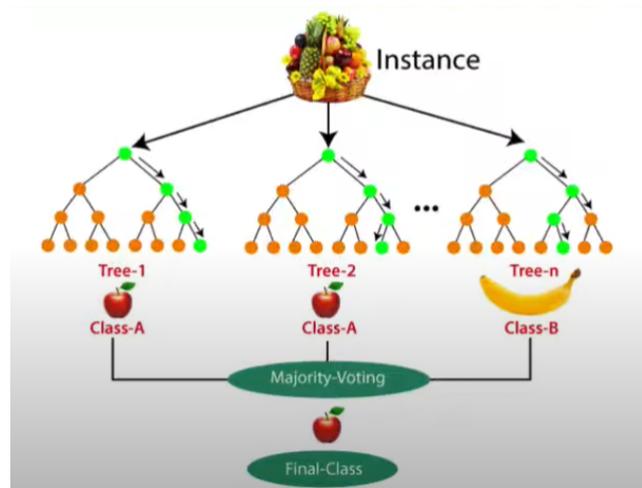


Figura 1: Diagrama ilustrando o funcionamento do modelo de Floresta Aleatória (PROFDANILO_DS, 2021).

Avaliação

O desempenho do modelo foi avaliado utilizando três métricas principais: RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) e R^2 (Coeficiente de Determinação), que forneceram uma visão clara sobre a precisão e robustez do modelo. Os resultados das comparações confirmaram a superioridade da Floresta Aleatória em termos de precisão e capacidade de generalização.

Após a avaliação, o modelo foi salvo e exportado para permitir sua reutilização em futuros estudos ou aplicações práticas, assegurando sua compatibilidade com diferentes ambientes de produção. A metodologia adotada contribui significativamente para o avanço da pesquisa em previsão energética em cenários urbanos complexos.

RESULTADOS E DISCUSSÃO

Este estudo teve como objetivo avaliar a eficácia do modelo de Floresta Aleatória na previsão de consumo energético em três zonas distintas da cidade de Tetuão, Marrocos. Os resultados foram comparados com outros modelos de aprendizado de máquina, incluindo Regressão Linear, Regressão Ridge, Regressão Lasso, K-Nearest Neighbors (KNN), Support Vector Regressor (SVR), Gradient Boosting, AdaBoost, Extra Trees e XGBoost. As métricas de desempenho utilizadas foram RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) e R^2 (Coeficiente de Determinação).

Desempenho do Modelo de Floresta Aleatória

Os resultados para o modelo de Floresta Aleatória, apresentados na Figura 2, mostram que este modelo teve desempenho superior em todas as zonas analisadas, com RMSE e MAE significativamente menores e valores de R^2 próximos de 1,0. Estes resultados indicam uma forte correlação entre os valores previstos e os reais, destacando a precisão e a robustez do modelo.

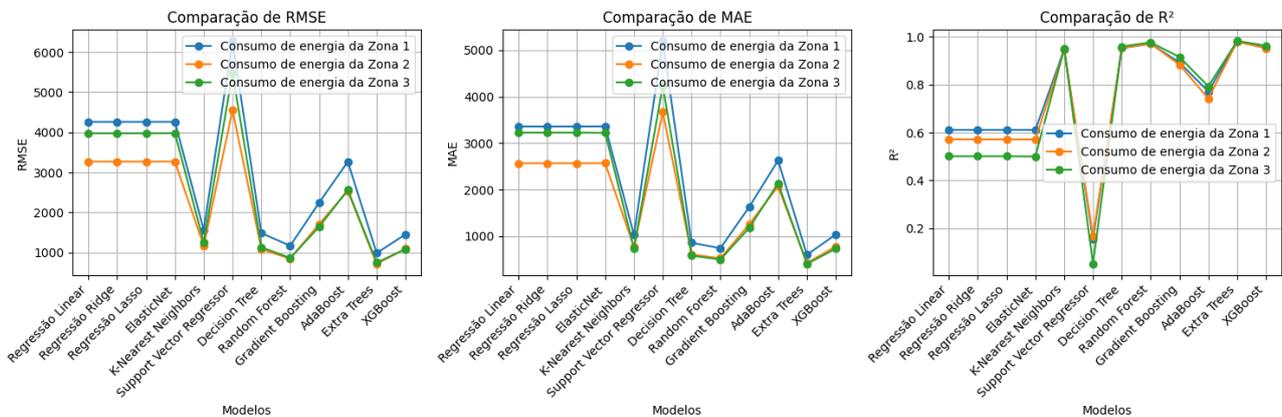


Figura 2: Comparação de Desempenho entre Diferentes Modelos

Zona 1: O modelo apresentou um RMSE de 1.147,21, um MAE de 722,00 e um R^2 de 0,9718, demonstrando alta precisão.

Zona 2: Nesta zona, os valores de RMSE e MAE foram 842,59 e 523,70, respectivamente, com R^2 de 0,9715.

Zona 3: A Zona 3 apresentou os melhores resultados, com RMSE de 865,77, MAE de 489,66 e R^2 de 0,9763.

Comparação com Outros Modelos

A análise comparativa dos modelos (Figura 2) revela que a Floresta Aleatória supera os demais modelos, especialmente em termos de RMSE e MAE. O SVR (Support Vector Regression) apresentou o maior RMSE, demonstrando menor eficácia comparado à Floresta Aleatória.

Modelos Baseados em Árvores: O Gradient Boosting e o XGBoost tiveram desempenhos próximos, mas inferiores em precisão, sugerindo que a Floresta Aleatória é mais adequada para este contexto.

Modelos Lineares: A Regressão Linear, Ridge e Lasso foram menos eficazes, com RMSE e MAE maiores, indicando dificuldade em capturar a complexidade dos dados.

Discussão dos Resultados

Os resultados indicam que o modelo de Floresta Aleatória se destacou na previsão do consumo energético, mostrando robustez significativa em diferentes zonas urbanas de Tetuão, Marrocos. A variabilidade nos resultados pode ser atribuída às diferenças demográficas e climáticas que influenciam os padrões de consumo nas zonas estudadas. A superioridade deste modelo em relação a abordagens mais simples, como a Regressão Linear, evidencia sua capacidade de capturar a complexidade inerente aos dados urbanos.

A análise de *outliers*, refletida no valor de RMSE, sugere que a Floresta Aleatória lida eficazmente com dados ruidosos, o que é essencial para aplicações práticas. Além disso, a combinação de variáveis climáticas e temporais aprimorou a precisão das previsões, validando a hipótese central do estudo.

Comparação com Outros Estudos

Para contextualizar e validar os resultados, as métricas de desempenho foram comparadas com outros estudos que utilizaram diferentes algoritmos para a previsão de consumo energético. Este estudo, com um MAE de 578.45, RMSE de 951.19 e R^2 de 0.97, destacou-se em termos de precisão quando comparado a outros estudos, como o de (ZOGAAN, 2022), sugerindo que o modelo foi altamente ajustado para o problema específico abordado.

Ao comparar com outros classificadores, como o Random Forest utilizado por (SÉRIO, 2020), que obteve um RMSE de 2223.0 e um R^2 de 0.97, é possível inferir que melhorias significativas foram implementadas no modelo proposto neste estudo. Outros modelos, como a Rede Neural de (SALAM; HIBAOU, 2018) e o KNN de (SHAPI; RAMLI; AWALIN, 2021), apresentaram maior MAE, sugerindo menor precisão em relação ao Random Forest.

Modelos mais avançados, como o Parallel ConvLSTM (KSHETRIMAYUM; SINGH; HOQUE, 2023), mostraram bom desempenho, com um R^2 de 0.98. No entanto, a simplicidade e a robustez do Random Forest proposto neste estudo tornam-no uma escolha mais prática e eficiente, especialmente considerando sua capacidade de lidar com dados ruidosos e evitar o sobreajuste.

A comparação com esses estudos valida a eficácia da Floresta Aleatória, destacando-se como uma solução robusta e precisa para previsões de consumo energético em ambientes urbanos complexos.

CONCLUSÕES

Este estudo demonstrou que o modelo de Floresta Aleatória é altamente eficaz na previsão de consumo energético em ambientes urbanos, utilizando a cidade de Tetuão, Marrocos, como estudo de caso. As comparações realizadas com outros modelos de aprendizado de máquina, como Regressão Linear e K-Nearest Neighbors, indicaram que a Floresta Aleatória oferece maior precisão e robustez, especialmente em contextos de alta variabilidade de dados.

A inclusão de variáveis climáticas e temporais foi essencial para melhorar a precisão das previsões, validando a hipótese de que essas variáveis são fundamentais para modelar padrões de consumo energético urbano. O elevado coeficiente de determinação (R^2) em todas as zonas analisadas reforça a capacidade do modelo em capturar a complexidade dos dados.

Assim, este trabalho não só avança o conhecimento na área de previsão de consumo energético, mas também oferece uma base prática e confiável para a aplicação de modelos preditivos em contextos urbanos. As implicações diretas incluem a melhoria na gestão de recursos e o planejamento de infraestrutura em cidades inteligentes, contribuindo para o desenvolvimento urbano sustentável.

CONTRIBUIÇÕES DOS AUTORES

Murilo Vargas da Silva e Vinícius de Souza Santos contribuíram com a concepção e escopo do estudo. Vinícius de Souza Santos procedeu com o desenvolvimento e escreveu o trabalho.

Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2024/02900-3.

Os autores também agradecem ao Departamento de Informática do IFSP, Câmpus Birigui, pela infraestrutura e apoio fornecidos durante a realização desta pesquisa.

REFERÊNCIAS

AMARAL, H. L. M. d. *Desenvolvimento de uma nova metodologia para previsão do consumo de energia elétrica de curto prazo*. Tese (Doutorado) — Universidade de São Paulo, 2019.

BARTOLO, R. d. S. D. S. *Previsão Inteligente do consumo de energia com base em modelos Deep Learning*. Tese (Doutorado) — Instituto Politecnico do Porto (Portugal), 2018.

BREIMAN, L. Random forests. *Machine learning*, v. 45, p. 5–32, 2001.

EPE. *Carga de energia deve crescer em média 3,4% por ano no período de 2022 a 2026*. 2023. Disponível em: <<https://www.epe.gov.br/pt/imprensa/noticias/carga-de-energia-deve-crescer-em-media-3-4-por-ano-no-periodo-de-2022-a-2026>>. Acesso em: 9 jul. 2024.

GODINHO, X. D. V. *Previsão de consumos de energia associados à climatização em edifícios de serviços com base em estratégias de Machine Learning*. Tese (Doutorado) — Escola Superior de Tecnologia e Gestão de Leiria, 2020.

KSHETRIMAYUM, N.; SINGH, K. R.; HOQUE, N. P-convlstm: An effective parallel convlstm-based model for short-term electricity load forecasting. 2023.

NALLATHAMBI, S.; RAMASAMY, K. Prediction of electricity consumption based on dt and rf: An application on usa country power consumption. In: IEEE. *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*. [S.l.], 2017. p. 1–7.

PROFDANILO_DS. *Modelo Random Forest - Teoria*. 2021. YouTube video. Acessado em: 29 de agosto de 2024. Disponível em: <<https://www.youtube.com/watch?v=eQep5PbvEXg>>.

SALAM, A.; HIBAOUI, A. E. Comparison of machine learning algorithms for the power consumption prediction:-case study of tetouan city-. In: IEEE. [S.l.], 2018.

SANTOS, J. A. A.; CHAUCOSKI, Y. Previsão do consumo de energia elétrica na região sudeste: Um estudo de caso usando sarima e lstm. *Revista Cereus*, v. 12, n. 4, p. 93–104, 2020.

SHAPI, M. K. M.; RAMLI, N. A.; AWALIN, L. J. Energy consumption prediction by using machine learning for smart building. *Developments in the Built Environment*, v. 5, p. 100037, 2021.

SÉRIO, F. A. A. *Modelos de Machine Learning na Gestão de Consumos de Energia*. Tese (Doutorado) — Instituto Politecnico de Viseu (Portugal), 2020.

ZOGAAN, W. A. Power consumption prediction using random forest model. *International Journal of Mechanical Engineering*, Jazan, 45142, Saudi Arabia, v. 7, n. Special Issue 5, April-May, 2022. ISSN 0974-5823.