

PROPUESTA DE MONOGRAFÍA

Título del proyecto	Estimación del consumo de energía a nivel mundial utilizando técnicas de aprendizaje de máquina.	
Estudiante 1		
Nombres completos	Lina María Beltrán Durango	E-mail: linam.beltran@udea.edu.co
		GitHub: https://github.com/lmbd92/DataScienceMonograph
Estudiante 2		
Nombres completos	Mario Earles Otero Andrade	E-mail: mario.otero@udea.edu.co
		GitHub: https://github.com/lmbd92/DataScienceMonograph

1. Descripción del problema

La predicción del consumo de energía es un problema en el sector energético, ya que permite a las empresas y gobiernos planificar mejor la producción y distribución de energía. Además, también ayuda a evaluar el impacto ambiental del consumo de energía en diferentes regiones del mundo (Jung et al., 2021; Njuguna et al., 2020). El uso de técnicas de aprendizaje de máquina puede proporcionar una estimación precisa del consumo de energía en diferentes regiones del mundo. Esto se puede lograr utilizando datos históricos y actuales de consumo de energía, así como de otros factores relevantes como la población, el clima y el crecimiento económico (Zhao et al., 2020; Tian et al., 2021). En esta monografía, se llevará a cabo la estimación del consumo de energía a nivel mundial utilizando técnicas de aprendizaje de máquina (Girardello et al., 2019; Xu et al., 2020).

2. Planteamiento del problema

El consumo de energía es un tema crucial en la agenda global debido a su relación con el desarrollo económico y la sostenibilidad ambiental (Wang et al., 2020; Singh et al., 2021). La predicción del consumo de energía a nivel mundial es un problema complejo que requiere el uso de técnicas avanzadas de análisis de datos y modelos de aprendizaje de máquina (Huang et al., 2020; Zhang et al., 2021). A través de estas técnicas, es posible analizar una amplia variedad de factores que influyen en el consumo de energía, como la población, el clima, la actividad económica y los avances tecnológicos (Geng et al., 2020; Zhang et al., 2020). En esta monografía, se abordará el problema de la estimación del consumo de energía a nivel mundial utilizando técnicas de aprendizaje de máquina, con el objetivo de proporcionar una herramienta útil para la planificación energética y la toma de decisiones.

3. Estado del Arte

La estimación del consumo de energía a nivel mundial es un problema complejo que requiere una gran cantidad de datos y técnicas avanzadas de análisis. Algunos estudios recientes han utilizado técnicas de aprendizaje de máquina para analizar el consumo de

energía en diferentes regiones del mundo, como el estudio de Li et al. (2021) que utilizó un modelo de redes neuronales para predecir el consumo de energía en China. Otros estudios se han centrado en la estimación del consumo de energía a nivel mundial utilizando modelos de regresión y análisis estadísticos, como el estudio de Zhang et al. (2020) que utilizó un modelo de regresión para predecir el consumo de energía en todo el mundo. Estos estudios demuestran la importancia del uso de técnicas de aprendizaje de máquina en la estimación del consumo de energía a nivel mundial, lo que puede ayudar a los responsables de la toma de decisiones a tomar medidas para reducir el consumo de energía y promover el desarrollo sostenible.

4. Descripción del dataset

Este dataset se llama “Energy Dataset Country-Wise (1900-2021)” es información recopilada de: Our world in Data BP Statistical Review of World Energy SHIFT Data Portal Ember – Data Explorer and the Ember European Electricity Review. El dataset está compuesto por 241 archivos .csv que contienen información de 17239 registros que tienen 128 características de 241 países alrededor del mundo.

Luego de realizar un análisis exploratorio del dataset se decidió tomar información de 120 características que están presentes en los registros de 91 países ubicados en todos los continentes, en el lapso de tiempo de 1980 a 2021. El criterio de refinamiento del dataset fueron los datos faltantes, por ende, esta decisión es tomada con el fin de disminuir la cantidad de datos faltantes que va a tener el dataset con el que se va a trabajar, puesto que el dataset completo tiene un porcentaje de datos faltantes del 64%, y la selección de información realizada tiene un porcentaje de datos faltantes del 18%.

El link en donde se encuentra cargado el dataset es el siguiente:

<https://www.kaggle.com/datasets/pranjalverma08/energy-dataset-countrywise-19002021>

5. Métricas de desempeño

La selección de las métricas de desempeño adecuadas es fundamental para evaluar la precisión del modelo y garantizar que se estén tomando decisiones informadas en el sector energético. Según Koo et al. (2021), el error cuadrático medio (MSE) y el coeficiente de determinación (R^2) son las métricas más utilizadas para evaluar los modelos de regresión en la estimación del consumo de energía. Además, el error absoluto medio (MAE) y el error absoluto porcentual medio (MAPE) también son importantes para evaluar el desempeño del modelo y proporcionar una comprensión más completa de los errores de predicción. Adicionalmente, la validación cruzada es una técnica importante para evaluar la capacidad del modelo para generalizar datos nuevos y no vistos. Según Pérez et al. (2021), la validación cruzada permite evaluar el desempeño del modelo en diferentes conjuntos de datos y reducir el riesgo de overfitting o underfitting del modelo. Además, la validación cruzada es útil para determinar la robustez del modelo y proporcionar una evaluación más precisa del desempeño del modelo.

6. Criterios de desempeño

Los criterios de desempeño para los modelos de regresión serán la precisión, la

simplicidad, la estabilidad y la capacidad de generalización. La precisión se evaluará mediante las métricas de desempeño mencionadas anteriormente (Alfaware, 2021; Brownlee, 2018; Géron, 2019). La simplicidad se evaluará en función de la complejidad del modelo y la facilidad de interpretación, puede evaluarse utilizando métodos como la eliminación de características o la regularización (Géron, 2019). La estabilidad se evaluará mediante la sensibilidad del modelo a cambios en los datos de entrada, esta puede evaluarse mediante métodos como la validación cruzada repetida (Brownlee, 2018). Finalmente, la capacidad de generalización del modelo puede evaluarse mediante técnicas de validación cruzada, como la validación cruzada k-fold y la validación cruzada leave-one-out (Géron, 2019). La capacidad de generalización se evaluará utilizando técnicas de validación cruzada.

7. Referencias bibliográficas

- Alfaware. (2021). Regression Metrics: Mean Squared Error, R2 Score, Mean Absolute Error, Mean Absolute Percentage Error. Recuperado el 14 de abril de 2023, de <https://alfaware.com/machine-learning-regression-metrics-mean-squared-error-r2-score-mean-absolute-error-mean-absolute-percentage-error/>
- Brownlee, J. (2018). How to Choose a Feature Selection Method For Machine Learning. Recuperado el 14 de abril de 2023, de <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.
- Huang, L., Chen, Y., Wang, H., & Chen, B. (2020). A new hybrid method for short-term electricity consumption forecasting with uncertainty quantification. *Applied Energy*, 276, 115495.
- Islam, M. S., Islam, M. M., Rumi, R. K., & Ahamed, S. I. (2019). Analysis of Solar Energy Generation Using Machine Learning Techniques. In 2019 IEEE International Conference on Power, Instrumentation, Control and Computing (PICC) (pp. 1-5). IEEE. doi: 10.1109/PICC47502.2019.9093975
- Jain, A., Kumar, A., & Kumar, A. (2020). A Comparative Study of Machine Learning Techniques for Electricity Consumption Forecasting. *IEEE Access*, 8, 190117-190128. doi: 10.1109/ACCESS.2020.3030683
- Jia, R., Zhang, Y., & Zheng, C. (2021). Improved Hybrid LSTM Model for Energy Consumption Forecasting. *IEEE Transactions on Industrial Informatics*, 17(2), 1107-1116. doi: 10.1109/TII.2020.2991871
- Jung, J. Y., Shin, Y. H., & Park, K. Y. (2021). Forecasting residential electricity consumption using machine learning and energy big data analysis. *Energies*, 14(2), 316.
- Li, W., Liu, L., Liu, Y., & Li, R. (2021). Short-Term Forecasting Model of Power Load Based on a Combined Neural Network Algorithm. *IEEE Access*, 9, 127454-127467.
- Njuguna, M., Kamangira, K., & Mvungi, N. H. (2020). Electricity demand forecasting using machine learning: A review. *Energy Reports*, 6, 974-983.

Singh, P. K., Gupta, R., & Tiwari, A. K. (2021). A review on energy consumption, economic growth, and environmental degradation nexus. *Journal of Cleaner Production*, 278, 123822.

Tian, Y., Xu, Y., & Shen, Y. (2021). Electricity demand forecasting for urban residential sector using machine learning: A case study of Shanghai, China. *Energy*, 223, 119774.

Wang, H., Zhang, H., Yang, X., & Wang, X. (2020). Analysis of the relationship between energy consumption and economic growth: A comparative study of developed and developing countries. *Energy Reports*, 6, 811-821.

Xu, X., Lu, Z., Wu, J., & Chen, S. (2020). A comparative study of machine learning techniques for energy consumption prediction. *IEEE Access*, 8, 203490-203498.

Zhao, X., Zhang, Y., Wang, Y., & Guo, X. (2020). Short-term electricity consumption forecasting using machine learning algorithms: A comparative study. *IEEE Access*, 8, 125254-125263.

Zhang, G., Zhang, X., & Han, Y. (2020). Short-term power load forecasting using machine learning and deep learning algorithms: A comparative study. *Energies*, 13(5), 1205.

Zhang, J., Chen, Y., & Zhang, Y. (2020). Estimating Global Energy Consumption Using a Hybrid Regression Model. *Energy Procedia*, 158, 722-727.

Zhang, Y., Wei, Y. M., Che, X., & Liang, Q. (2021). Big data, machine learning, and smart energy management: A review and future research directions. *Applied Energy*, 280, 116008.