EGU24-4116, updated on 07 Feb 2025
https://doi.org/10.5194/egusphere-egu24-4116
EGU General Assembly 2024

Stream Machine Learning for Lightning Nowcasting - Harnessing the Power of Continuously Updated Data

Cesar Beneti[1], **Luis Pavam**[1,2], Luiz Oliveira[2], Marco Alves[2], Leonardo Calvetti[3], and Fernanda Verdelho[1]

[1]SIMEPAR - Parana Environmental Technology and Monitoring System, Curitiba, Brazil (cesar.beneti@simepar.br)
[2]UFPR - Parana Federal University, Department of Informatics, Curitiba, Brazil
[3]UFPEL - Federal University of Pelotas, Pelotas, Brazil

Uninterrupted access to electricity is a fundamental feature of civilization. In its absence, an all-embracing cessation of activities occurs, ranging from essential services to more frivolous activities. The maintenance of the energy supply is critical for society's day-to-day functions. The Brazilian state of Paraná (PR) is home to the world's second-largest hydropower plant, Itaipu, which, in conjunction with other power plants in the state, provides almost one-third of the power energy production in Brazil. The transmission lines that pervade PR are essential to Brazil's power distribution system, for hydropower generation is typically made far away from the regions that most demand it, being transported by transmission lines in an interconnected power grid. This type of asset mainly depends on the forecast of Cloud-to-Ground (CG) lightning, as it is one of the leading weather-related causes of power outages. Lightning and wind gusts are the two leading weather-related causes of disruptions, representing at least 23% of the known causes of energy disruption, as declared by the local power distribution company. Our study of lightning incidence and power outages from 2017-2021 indicates a correlation of 0.98 between these events, denoting that more outages must be lightning-related. Reliable CG lightning forecasts are crucial for proactive hazard mitigation. This work expounds on developing a Machine Learning (ML) model for CG lightning forecasting for PR. Our ML model predicts the occurrence or lack of CG lightning near power company assets in PR, defining a binary classification task. The model makes its predictions based on the past spatio-temporal conditions of lightning occurrences, requiring only past lightning data to forecast lightning. We chose to use a stream ML method, i.e., the model is continuously trained as new data arrives. Using a stream ML, we intend to harness the machine's capacity to continuously learn the patterns of lightning occurrence and power outages in real-time -- thus constructing an ever-updating model capable of adapting to transient weather conditions. Given its rapid training time and aptitude for classification tasks, the chosen algorithm was a Very Fast Decision Tree. The stream ML classifier outperforms a classic static ML model by 30% regarding the ROC AUC metric (stream: 71.80%, static: 40.85%) and 50% considering the Micro-f1 score (stream: 91.05%, static: 40.91%). These results arise from the highly dynamic nature of lightning, defining an ideal phenomenon for prediction based on a constantly updated stream of data. An automatic system for CG lightning forecasting for power company assets is helpful for risk management and operational planning. Future steps include increasing the lead time from ten

min. to up to one hour, allowing for more time to prepare and anticipate hazards, preventing power outages, and optimizing personnel allocation.