AI/ML FOR ELECTRIC DISTRIBUTION GRID OUTAGE ANALYIS

STUDENTS: Laxman Balamurugan, Shiyu Chen, Andrew Nguyen, Sreeram S, Louis Serchuk, Yifei Zhao, Kaiyuan Zheng

GE VERNOVA

Abstract

Traditionally, power outage management relies on preventive maintenance, robust infrastructure, and emergency responses. However, these methods often fail to prevent unexpected failures or adapt quickly to changing conditions. Integrating AI into power systems enhances these methods. AI and ML models analyze data from weather, grid usage, and equipment status to predict failures before they occur.

Project Requirements:
- Collect and preprocess outage and weather data
- Test and train the cleaned data using a wide variety of models
- Analyze results

Data Collection

Data was collected from three different sources:

- Power outage data was purchased from **PowerOutage.us**
- Weather data was collected from the **Nation Centers for Environmental Information (NCEI)** and **OpenWeatherMap** and combined to form a comprehensive database

PowerOutage.US OpenWeather National Centers for Environmental Information

Model Plan

Problem Objective: Develop a Machine Learning model that will predict the percentage of power outages in a Florida county based on outage and weather data

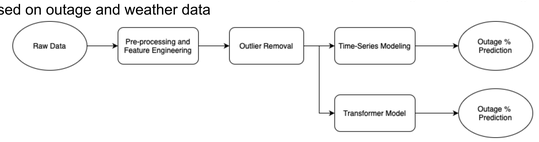

Figure 1: High level flowchart of design process

Data Processing

- Categorial features were **numerically encoded.**
- New feature of '**Holiday/Weekday**' was introduced to inspect the effect of public holidays and weekends on Outage % as electricity usage is highly impacted by this metric.
- Features that were **redundant** with each other were **redacted** using Correlation matrix selection.
- Features with Correlation value higher than **0.8** with the Target variable (Outage %) are retained.
- DBSCAN Clustering method was used to remove outliers from the dataset.

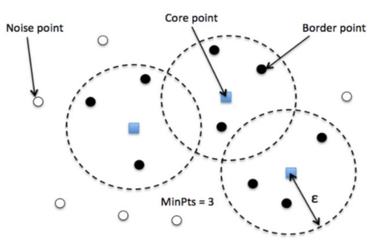

Figure 2: Visualization for how DBSCAN clustering method operates

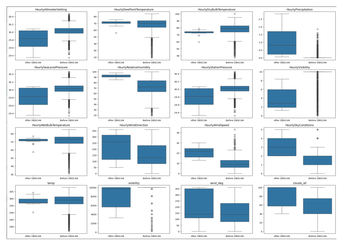

Figure 3: Box Plots of notable weather features showing the result of the DBSCAN outlier cleaning process

Successful Training Models

We followed two types of ML training pipeline: One is Time-Series modeling and Independent modeling. Our AI solution Roadmap for Power outage prediction is shown below:

Time Series Model

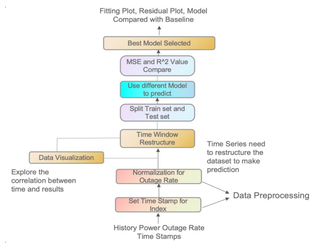

Figure 4: Time Series Model Flow Chart

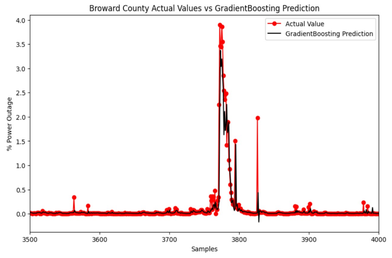

Figure 6: Actual percentage of power outages compared to the predicted power outages in Broward County according to the Time Series Model

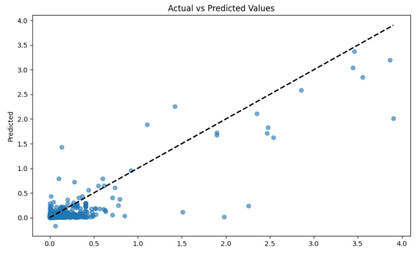

Figure 8: Regression Plot - Actual percentage of power outages vs the predicted power outages in Broward County according to the Time Series Model

Transformer Model

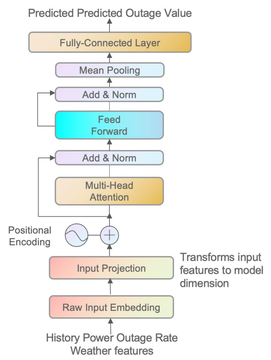

Figure 5: Transformer Model Flow Chart

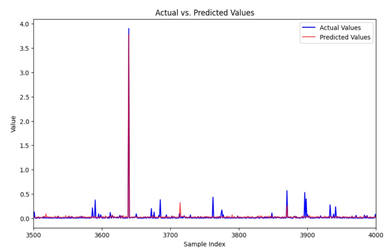

Figure 7: Actual percentage of power outages compared to the predicted power outages in Broward County according to the Transformer Model

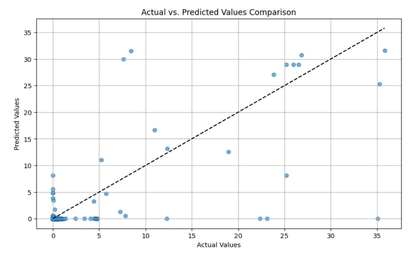

Figure 9: Regression Plot - Actual percentage of power outages vs the predicted power outages in Broward County according to the Transformer Model

Other Models Attempted

Before arriving at our Time-Series and Transformer based models, we experimented the Outage % prediction with foundational/trivial models like Linear Regression, SVM, Tree based methods, SVM, by following Occam's Razor, then proceeding with higher level models upon realizing the performance and limitations of the primary models attempted.

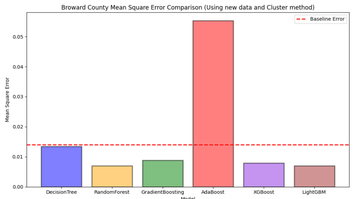

Figure 10: Mean Square Error comparison between the six primary models attempted compared to the baseline MSE

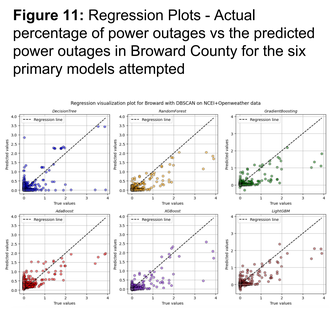

Figure 11: Regression Plots - Actual percentage of power outages vs the predicted power outages in Broward County for the six primary models attempted

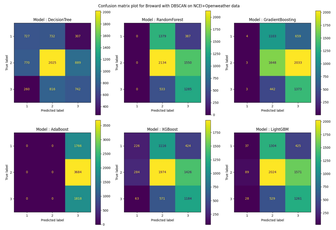

Figure 12: Confusion Matrices for the six primary models attempted for Broward County, by splitting into 3 classes from the Outage % predicted based on the 25th and 75th percentile of the outage values

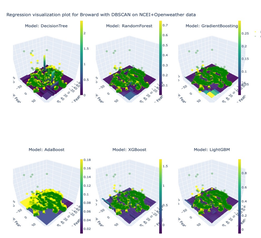

Figure 13: Regression Visualization plot by reducing the dimensionality of the features to 2 via t-SNE, to have surface level visualization idea on how the primary models predict outage %

Attempted Models Overview

MODEL	R^2	MSE
Time Series	0.7720	0.0065
Transformer	0.6406	0.0065
Decision Tree	0.2554	0.0134
Random Forest	0.6130	0.0070
Gradient Boosting	0.5104	0.0088
AdaBoost	-2.0706	0.0553
XGBoost	0.5628	0.0079
LightGBM	0.6133	0.0070

Figure 14: Summary of the main models tried that gave the most promising results for Broward County

Conclusion

- The goal for this project was to find the best fitting model for predicting power outages based on weather data.
- The best fitting model based on R^2 values is the Time Series Model.
- The next highest was the Transformer Model.

Future Work

Future work could include expanding to different states, trying more advanced methods, or collecting more data to try and predict where the outage occurs during transmission.

References and Acknowledgments

We would like to thank our sponsor GE Vernova and our advisors Armina Foroutan and Baosen Zhang for their help with our project!

Wang, Renfeng; Vanga, Venkata Leela 'MG'; Zaiken, Zachary B.; and Bennett, Jonathan (2022) "Analysis of the electric power outage data and prediction of electric power outage for major metropolitan areas in Texas using Machine Learning and Time Series Methods," *SMU Data Science Review*: Vol. 6: No. 1, Article 5.

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

ADVISERS: Baosen Zhang

SPONSOR: Armina Foroutan & GE Vernova