JOURNAL OF
EMERGING INVESTIGATORS

Article

Weather-based power outage prediction in New York City: An ensemble machine learning approach

Ashwin Mohan[1], **Madras Mohanasundaram**[2]
[1]Coppell High School, Coppell, Texas

SUMMARY

Power outages pose significant challenges in urban areas, even during non-extreme weather conditions. We hypothesized that variations in weather conditions (e.g., temperature, precipitation, humidity, snow, wind, etc.) would be predictive of power outages. Our study examined the relationship between weather variables and outage occurrences in New York City, using historical data and machine learning techniques. The ensemble regression model in this research consisted of random forest, extreme gradient boosting regressors, support vector regression, and multilayer perceptron base models with gradient boosting regressor meta model. The model achieved moderate accuracy on the most predictive dataset (mean absolute error = 3.774, mean absolute percentage error = 34.073%, mean squared error = 24.874, and R^2 = 0.824). Correlation analysis identified energy demand and temperature as the features most strongly associated with outage frequency, and that rolling averages tend to be better predictors than daily values. Principal component analysis revealed that days with extreme temperatures had more outages, and that temperature, humidity, and precipitation were the main drivers of variance. The limitations of this approach include unaccounted infrastructure factors, complexity of power failure causes, and timing discrepancies in outage reports. Exploring alternative modeling techniques and expanding the dataset to other geographic regions will further refine the findings. This analysis contributes to our understanding of how urban energy systems respond to climate variability and inform strategies for enhancing power grid resilience.

INTRODUCTION

Power outages are a recurring issue in urban areas. While extreme weather is frequently cited as the major cause of power disruptions, more gradual weather phenomena, including prolonged periods of high heat and precipitation are increasingly recognized as critical factors (1-3). The urban heat island (UHI) effect intensifies temperatures during summer months, leading to increased electricity demand as residents and businesses heavily rely on air conditioning systems for relief (4). This heightened demand can strain power grids, potentially resulting in outages that are often worsened by aging infrastructure or inadequate capacity (5).

Machine learning (ML) has gained significant attention in electric grid analysis in recent years. Many studies have demonstrated the ability of machine learning to forecast outages due to extreme weather events. For example, statistical and fragility-based models have been able to achieve high accuracy when predicting power outages caused by hurricanes and severe storms, though large amounts of data are often necessary (2). Deep learning approaches have been used to predict outages under multiple types of extreme conditions, achieving a low classification error of 0.022 (6). These models utilized historical weather data and grid performance metrics, highlighting the potential for machine learning in optimizing grid resilience. However, while much of the research has focused on extreme weather, studies that integrate machine learning with the specific effects of phenomena like UHI effects remain limited. While not an ML model, two-way fixed effects regression has been applied to forecast electricity consumption spikes in urban areas in China during heat waves, predicting demand surges that could lead to outages (3). Few studies have applied these technologies specifically to urban heat-related outages, as current models have not fully addressed the unique challenges of urban environments.

New York City's population density is approximately 27,469 people per square mile, contributing to the intensity of the UHI effect due to the absorption and retention of heat by man-made surfaces (7). The UHI effect is most pronounced during summer months, especially at night. A study by NASA's Goddard Institute for Space Studies found that variations in New York City's UHI are correlated to power consumption during summer (4). Understanding such factors is important for predicting outages and enhancing resilience.

While extensive research exists on the impact of severe weather, there are few studies focusing on predicting outages due to more regular, albeit impactful weather conditions like heat waves (1-3, 6). We sought to address this gap by exploring the relationship between weather patterns, specifically heat-related conditions and the frequency of power outages in New York City. We hypothesized that variations in weather conditions like temperature, precipitation, humidity, snow, wind, etc. would be predictive of power outages. We identified energy demand and temperature as the features most strongly associated with outage frequency, and that rolling averages tend to be better predictors than daily values. Our analysis also revealed that days with extreme temperatures had more outages, and that temperature, humidity, and precipitation were the main drivers of variance.

RESULTS

We employed an ensemble machine learning regression method to test our hypothesis that variations in weather conditions are predictive of power outages and to accurately predict weather-related power outages in New York City. We used random forest, extreme gradient boosting (XGBoost or XGR), support vector regression (SVR), and multilayer

https://doi.org/10.59720/25-197

Model	MAE	MAPE	MSE	R²
Random forest	4.51	42.43%	38.61	0.63
SVR	4.86	39.04%	50.45	0.34
XGBoost	4.51	41.42%	37.03	0.58
XGR	4.53	37.46%	43.84	0.46
MLP	4.29	34.39%	35.78	0.48
Ensemble (GBR)	3.77	34.07%	24.87	0.82

Table 1. Accuracy metrics of ML models on 2019-2024 weather and energy data. Ensemble model was more accurate than individual base models. MAE: mean absolute error, MAPE: mean absolute percentage error, MSE: mean squared error, R^2: R-squared value.

perceptron (MLP) base models, each offering unique strengths in capturing complex data patterns. Random forest is effective at handling non-linear relationships and resists overfitting. XGBoost is a highly efficient boosting algorithm that optimizes regression models by building decision trees in a sequence, where later trees correct errors of previous ones. SVR uses a hyperplane to form predictions, effective for both linear and non-linear relationships and useful for managing high-dimensional data with reduced overfitting. MLP is a type of neural network model that learns complex patterns by adjusting internal weights and is flexible to wide ranges of data types. MLP builds on prediction gaps that may be left by the other simpler models. We trained each base model independently and used their predictions as features for the gradient boosting regressor (GBR) meta-model. We used the GBR to optimally combine the base models' outputs to produce the final prediction, improving accuracy by compensating for individual model weaknesses **(Table 1)**.

We trained and tested the model on three datasets – the first containing weather data from 2014 to 2024, the second from 2019 to 2024, and the last also from 2019 to 2024 but containing energy demand and energy generation data in addition to weather data. The last dataset helped contextualize the power outage data by providing insights into electricity consumption and production patterns. We evaluated the accuracy of the machine learning model using mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R-squared value (R^2). The model predicted power outages with moderate accuracy: MAE = 3.774, MAPE = 34.073%, MSE = 24.874, and R^2 = 0.824 **(Table 2)**.

We performed hyperparameter tuning on each model in the ensemble using the scikit-learn random search method to optimize accuracy. We minimized the MAE metric of each model, and this generally led to improvements in the other accuracy metrics described above. In the random forest model, we tuned the following key parameters to optimal values: number of trees = 500, minimum number of samples for splitting = 3, and maximum features evaluated at each split = square root of total features. In the XGBoost model, we tuned the following key parameters to optimal values: number of trees = 100, learning rate = 0.03, and tree depth = 5. In the SVR model, we tuned the following key parameters to optimal values: kernel type = polynomial, penalty parameter (C) = 1.0, and margin of tolerance (epsilon) = 4.45. In the MLP model, we tuned the following key parameters to optimal values: neurons per layer = 50, learning rate = 0.005, activation function = rectified linear unit (ReLU), and solver = limited-memory Broyden–Fletcher–Goldfarb–Shanno Algorithm (L-BFGS). We computed these metrics by repeatedly running the model for several iterations with different random splits of training and testing data. The model performed best when trained on the 2019–2024 dataset containing both weather conditions and energy usage variables, suggesting that including energy demand and generation improved predictive power. Comparing the predictions for the two weather-only datasets showed better accuracy for the full 2014–2024 dataset demonstrating the positive impact of larger training data.

Dataset	MAE	MAPE	MSE	R²
Weather (2014-2024)	4.124	36.299%	27.383	0.802
Weather (2019-2024)	4.343	39.854%	34.869	0.762
Weather and Energy (2019-2024)	3.774	34.073%	24.874	0.824

Table 2. Accuracy metrics of ensemble ML model with different training sets (10-12). Highest accuracy was achieved when using both weather and energy data. MAE: mean absolute error, MAPE: mean absolute percentage error, MSE: mean squared error, R^2: R-squared value

We used principal component analysis (PCA) to capture the directions of greatest variation in the data. In PCA, the first principal component (PC1) indicates the direction of maximum variance in the data and captures the single most informative axis. The second principal component (PC2) is orthogonal to PC1 and indicates the direction of the next-highest variance and captures a different, independent pattern in the data. In our analysis, PC1 explained 35.6% of the total variance and was primarily influenced by temperature-related variables, while PC2 was responsible for 13.8% of total variance and was tied to humidity and precipitation **(Figure 1)**. In the PCA score plot visualization, high-outage days tended to follow low and especially high magnitudes of PC1, suggesting a strong relationship between temperature extremes and power failures **(Figure 2)**. This visualization also showed a weaker association with PC2, though some high-outage days appeared toward its upper range.

We used Pearson correlation analysis to better

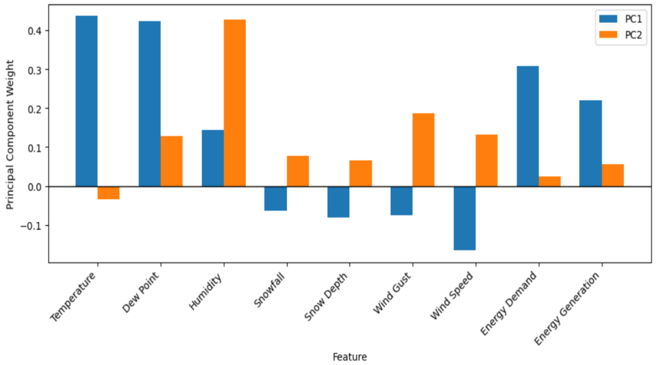

Figure 1. Principal component weights of features. Temperature, dew point, and energy variables contribute heavily to PC1 while humidity is the main influence on PC2.

https://doi.org/10.59720/25-197

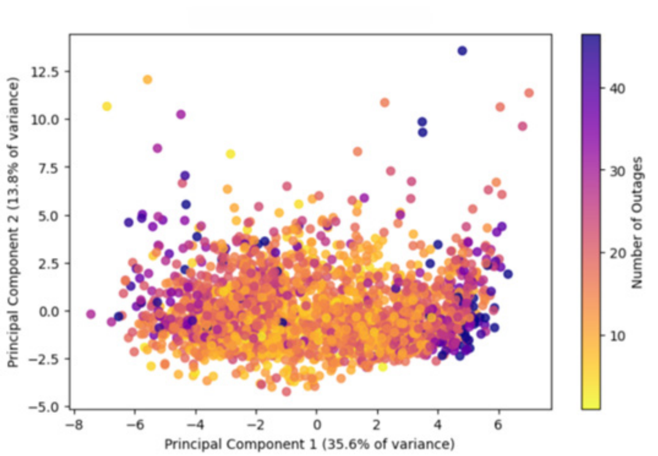

Figure 2. PCA score plot with principal components and number of outages. The yellow/red dots indicate lower number of outages while blue/purple dots indicate higher number of outages.

understand which variables were most associated with power outages **(Figure 3)**. Energy demand, energy generation, and temperature deviation from average (z-score) were the strongest predictors of outages. For example, the top four correlation coefficients were 0.457, 0.453, 0.407, and 0.152 for energy demand, energy generation, temperature deviation, and seven-day rolling average precipitation. The temperature z-scores were calculated as number of standard deviations from the mean and reflected how far daily temperatures deviated from typical values **(Figure 4)**. For example, 15.7°C, 31.9°C, and -12.5°C corresponded to z-scores of 0.20, 1.99, and 2.92, respectively. Greater temperature deviations were linked with more frequent power failures. In general, rolling averages for weather features correlated more strongly than raw daily values with the number of outages. For example, the correlation coefficients for daily precipitation, three-day rolling precipitation, and seven-day rolling precipitation were 0.091, 0.126, and 0.152.

DISCUSSION

We hypothesized that variations in weather conditions (e.g., temperature, precipitation, humidity, snow, wind, etc.) would be predictive of power outages. We examined the relationship between weather variables and outage occurrences in New York City, using historical data and machine learning techniques. The ensemble machine learning model predicted power outages with moderate accuracy **(Table 2)**. We used PCA to reduce dimensionality without losing the most critical variance in the data. We analyzed PCA loadings to determine which variables contributed the most to each principal component **(Figure 1)**. This analysis showed that temperature contributed heavily to PC1 while humidity was the main influence on PC2. The PCA score plot visualization mapped each data point based on the two principal components **(Figure 2)**. We observed higher number of outages at the low and especially high end of the PC1 range suggesting a strong relationship between temperature extremes and power failures. The visualization showed a weaker relationship between PC2 (humidity) and outage occurrences.

To better understand which variables were most associated with power outages, we conducted a Pearson correlation analysis. The resulting correlation coefficients reflect the strength of linear relationships between individual features and the number of daily outages **(Figure 3)**. It showed that energy demand has the strongest correlation with outages, reflecting the fact that most failures occur during high usage intervals that cause grid stress (10). The next strongest variable was electricity generation, which is also closely tied to demand, as utilities actively scale output to match consumption in real time (12). The third strongest feature was temperature deviation from the average as measured by z-scores. This correlates with heavy use of HVAC systems on very hot or cold days. Pronounced heat and cold also directly affect electrical infrastructure—high temperatures can lead to equipment overheating and reduced transmission efficiency, while extreme cold can make power lines brittle and increase mechanical failure risks (2, 5). After temperature, we found that other weather-related features such as precipitation, dew, snow, wind, etc. were weakly correlated with outages. Among weather variables, the highest correlations are mostly found in seven-day rolling averages, suggesting that sustained weather patterns are more relevant to outage occurrence.

While this model performed well in identifying key predictors of power outages, there are several limitations to consider. The outage complaint data may include delays between the outage time and the reported time, and in some cases, outages may not be reported at all. This is especially likely in poorer communities with lower participation in city services, which introduces potential bias into the data (10). In addition, the weather and energy features are based on daily city-wide values, which may overlook short-term or localized conditions that contribute to outages but are not captured in the aggregated data. Despite the professional nature of data collection, sensor inaccuracies could also contribute to error (8).

Using daily averages for weather and energy demand is unlikely to capture sharp fluctuations, such as sudden spikes in temperature, wind, or electricity usage that can

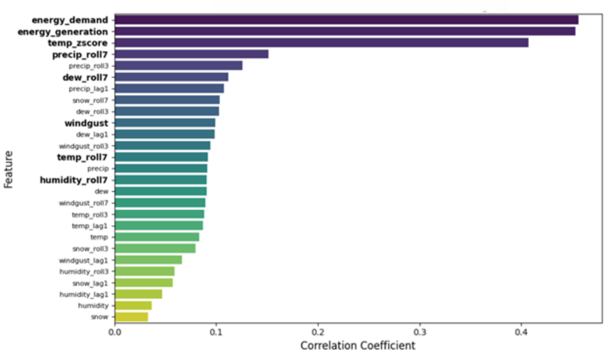

Figure 3. Correlation of features with number of outages. Correlation coefficient values were determined by Pearson correlation analysis of feature variables with number of outages. Yellow/green colors represent lower correlation and blue/purple colors represent higher correlation. Rolling averages for weather features (denoted in feature name with the ending _roll3 and _roll7) correlate more strongly than raw daily values. The most strongly correlated features in each category like energy demand, energy generation, temperature, precipitation, dew, wind, and humidity are labeled in bold.

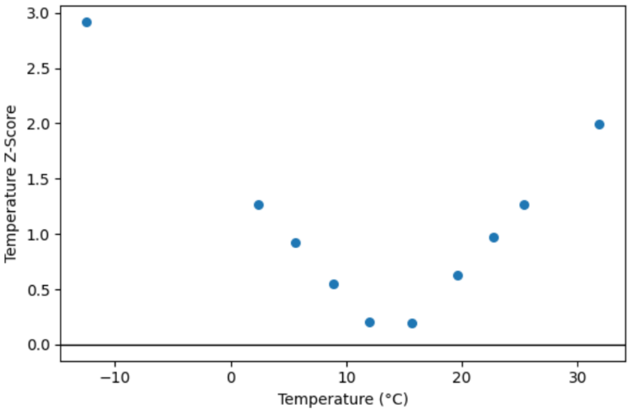

Figure 4. Z-scores for temperature deviations. Greater temperature deviations were linked with more frequent power failures.

place unexpected strain on the grid. Incorporating higher-frequency data, such as hourly weather or load information, could provide a clearer view of how specific events contribute to disruptions. The model also does not account for any infrastructure-related variables, such as the age or condition of equipment, maintenance records, or grid layout. Including this kind of operational data could help build a more complete picture of outage risk (8).

Lastly, although the stacking ensemble method was selected for improved overall accuracy, it may be difficult to scale and/or implement in real-time settings due to the computational complexity (8, 9). For practical applications, particularly in real-time forecasting or across larger geographic areas, model simplification or algorithmic optimization may be necessary. Future work could also explore integrating spatial modeling techniques, such as geospatial clustering or graph-based representations of the power grid, to enhance predictive power (8). Further investigations could explore the use of localized real-time weather data. Additionally, expanding the model to incorporate other forms of power system stress, such as aging infrastructure, could provide a broader view for predicting and preventing outages.

Feature engineering techniques, especially rolling averages, improved predictive accuracy by capturing short-term trends and deviations from typical conditions. Correlation analysis revealed that energy demand was the strongest individual predictor of outages and that large departures from the average daily temperature—both hot and cold—are also key indicators of grid stress. PCA further supported these trends, showing that days with very high or low temperatures tended to be associated with higher outage counts. These findings can help inform urban planners and infrastructure developers by identifying the factors that make regions within a power grid more vulnerable.

MATERIALS AND METHODS

Datasets

We utilized two primary datasets to investigate the relationship between weather conditions and power outages in New York City. The first dataset consisted of power outages reported by residents through the city's official portal from March 2014 to January 2024 **(Figure 5)**. This dataset, made publicly available on the New York City government website, provides detailed logs of individual power outage reports (10). Each entry includes information such as the date and location of the outage, as well as additional context provided by the reporting residents. To match the frequency of the explanatory data, the dataset was restructured to consolidate daily reports, with one row per day representing the total number of outages reported on that day. Outliers were removed using standard statistical thresholds. Specifically, values lying outside 1.5 times the interquartile range (IQR) from the first or third quartile were excluded. For example, in the daily average temperature data, values below approximately −10°C or above 40°C were considered outliers, as such extremes are rare in New York City's climate and likely reflect instrument or reporting errors rather than realistic weather observations.

The second dataset contains daily weather observations for New York City, covering the same time period. This dataset includes variables such as daily high and low temperatures, precipitation amounts, solar radiation, etc. This dataset is publicly accessible via the Visual Crossing Weather platform and was collected by professional weather instruments, ensuring the accuracy and reliability of the recorded measurements (11). Additionally, U.S. Energy Information Administration (EIA) supplementary data was incorporated to assess energy demand and generation trends in New York City (12).

These datasets were preprocessed using pandas and scikit-learn Python packages to convert date and non-numeric data for merging and numerical analysis by the machine learning models (13-15). Missing data points were present only in the wind gust top speed column, where they were substituted with estimates based on the average wind speed. To enhance the datasets, additional weather-related features were engineered. These include one-day lag variables, which capture the previous day's weather conditions, and three-day and seven-day rolling averages that represent short-term weather trends. These rolling values help smooth daily fluctuations and highlight sustained patterns such as ongoing heat or cold spells. A z-score for daily temperature was also calculated to indicate how much a day's temperature deviates from the recent mean, helping the model identify days that were pronouncedly hot or cold.

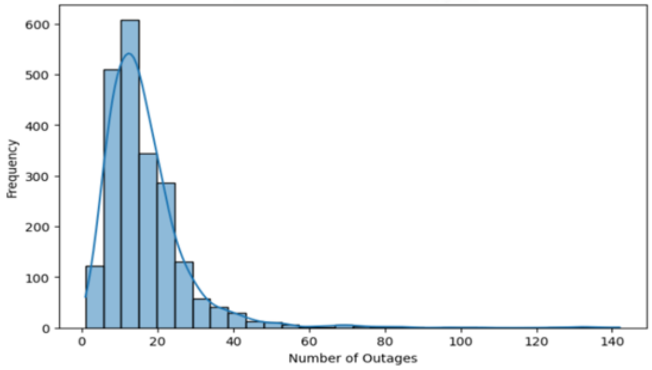

Figure 5. Histogram of number of daily power outages in New York City. There are few outages on most days. There are a few days with high number of outages possibly due to severe weather, infrastructure issues, etc.

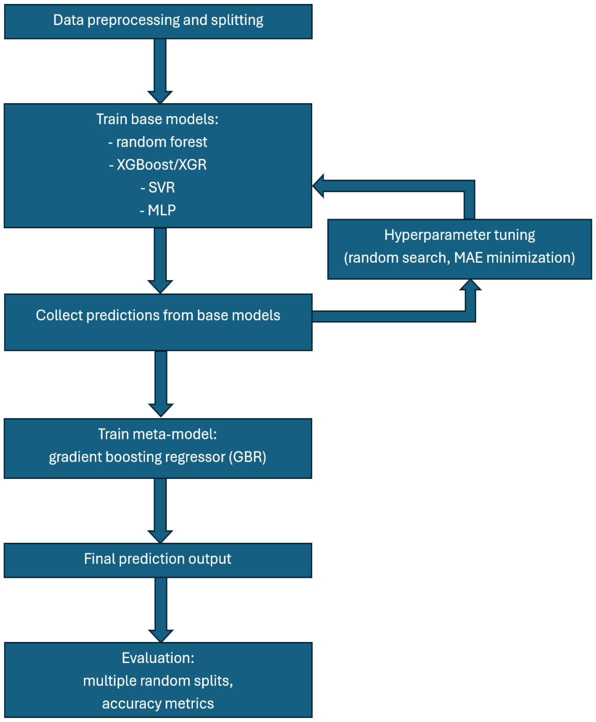

Figure 6: Machine learning pipeline flow-chart. Data was first preprocessed then base models were trained and tuned iteratively using the training data. The base models were then used to train the meta-model whose predictions were finally evaluated for accuracy metrics.

Three processed datasets of explanatory data were created for comparison. One contained 3814 observations of weather data from the full timeframe of 2014-2024. The other two contained 2188 observations from 2019 to 2024, with one containing just weather data and the other containing weather, energy demand, and energy generation data.

Machine Learning

To predict heat-related power outages in New York City, a stacking ensemble method which combined multiple machine learning models was employed to improve prediction accuracy **(Figure 6)**. In this approach, several base models were trained independently. The GBR meta-model combined the predictions of the base models to output the final prediction. This method reduced overfitting and bias by leveraging the strengths of different models **(Table 1)**. The machine learning pipeline was implemented in the Python programming language using the pandas and scikit-learn packages (13-15). The base models used in the ensemble were random forest, XGBoost or XGR, SVR, and MLP. Hyperparameter tuning was performed for each model in the ensemble through the scikit-learn random search method to optimize accuracy. The method minimized the mean absolute error metric of each model, and this generally led to improvements in the other accuracy metrics. These metrics were computed by repeatedly running the model for several iterations with different random splits of training and testing data.

Feature Analysis

Before model training, exploratory feature analysis was conducted using PCA and Pearson correlation to identify dominant predictors and reduce redundancy in the data. All continuous variables were standardized using z-score normalization to ensure that features with larger numeric ranges did not dominate the PCA results. PCA was then applied to the standardized dataset to identify the linear combinations of variables (principal components) that captured the most variance in the data. Pearson correlation analysis was also performed to measure the linear relationships between each explanatory variable and the number of outages. Plotting and visualization for feature analysis were done using the matplotlib and seaborn Python packages (16, 17).

Received: August 2, 2025
Accepted: November 19, 2025
Published: April 29, 2026

REFERENCES

1. AlHaddad, U., et al. “Towards Sustainable Energy Grids:A Machine Learning-Based Ensemble Methods Approach for Outages Estimation in Extreme Weather Events.” *Sustainability*, vol. 15, no. 16, 21 August 2023. https://doi.org/10.3390/su151612622.
2. Fatima, K., and H. Shareef. “Dynamic Bayesian Network Model for Overhead Power Lines Affected by Hurricanes.” *Forecasting*, vol. 7, no. 1, 5 March 2025. https://doi.org/10.3390/forecast7010011
3. Liang, J., et al. “Impacts of heatwaves on electricity reliability: Evidence from power outage data in China.” *iScience*, vol. 28, no. 2, 21 February 2025. https://doi.org/10.1016/j.isci.2025.111855
4. Gaffin, S., et al. “Variations in New York city’s urban heat island strength over time and space." *Theoretical and Applied Climatology*, vol. 94, 21 March 2008, pp. 1-11. https://doi.org/10.1007/s00704-007-0368-3
5. "Analyzing the Urban Heat Island Effect." New York City Department of Environmental Protection. https://www.nyc.gov/assets/dep/downloads/pdf/environment/education/10-analyzing-urban-heat-island-effect.pdf. Accessed 3 April 2025.
6. Rastgoo, R., et al. “Extreme outage prediction in power systems using a new deep generative informer model.” International *Journal of Electrical Power & Energy Systems*, vol. 167, June 2025. https://doi.org/10.1016/j.ijepes.2025.110627
7. "Table 2: Population, Land Area, and Population Density by County, New York State - 2020." New York State Department of Health. https://www.health.ny.gov/statistics/vital_statistics/2020/table02.htm. Accessed 10 April 2025.
8. Strielkowski, W., et al. “Prospects and Challenges of the Machine Learning and Data-Driven Methods for the Predictive Analysis of Power Systems: A Review.” *Energies*, vol. 16, no. 10, 11 May 2023. https://doi.org/10.3390/en16104025
9. Zhang, S., et al. “A Critical Review of Data-Driven Transient Stability Assessment of Power Systems: Principles, Prospects and Challenges.” *Energies*, vol. 14, no. 21, 2

https://doi.org/10.59720/25-197

November 2021. https://doi.org/10.3390/en14217238
10. "Power Outage Complaints." New York City Open Data. https://data.cityofnewyork.us/Social-Services/power-outage-complaints/br6j-yp22/data_preview. Accessed 10 April 2025.
11. "Weather Query Builder." Visual Crossing. https://www.visualcrossing.com/weather-query-builder/. Accessed 4 April 2025.
12. "Electricity Region Data." U.S. Energy Information Administration. https://www.eia.gov/opendata/browser/electricity/rto/region-data. Accessed 4 April 2025.
13. "Welcome to Python." Python. https://www.python.org. Accessed 2 April 2025.
14. "pandas: Python Data Analysis Library." pandas. https://pandas.pydata.org. Accessed 2 April 2025
15. "sciKit-learn: Machine Learning in Python." scikit-learn. https://scikit-learn.org/stable/. Accessed 3 April 2025.
16. "matplotlib: Visualization with Python." matplotlib. https://matplotlib.org. Accessed 3 April 2025.
17. "seaborn: Statistical Data Visualization." seaborn. https://seaborn.pydata.org. Accessed 3 April 2025.