Grid Disruption Analysis: Visualizing U.S. Grid Outages and Outage Prediction using Advanced Machine Learning
CSE 6242 Final Report

Team 155: Krishna Aryal, Crystal Vandekerkhove, Jinesh Patel

1 Introduction

Every year, thousands of people are affected by power outages caused by natural disasters. One of the most pressing challenges for electric companies and researchers is developing an effective power restoration plan while efficiently scheduling and routing repair crews to restore power as quickly as possible. Another major challenge is predicting when power outages will occur and how long they will last. This project addresses these challenges by developing a visual analytics platform that integrates machine learning to predict outage durations, providing actionable insights for utilities and communities.

2 Problem Definition

The ability to accurately estimate power outage durations has significant implications for both utility companies and the broader community. This information serves as a valuable tool in enhancing the efficiency and effectiveness of recovery efforts following power disruptions. For utility companies, precise outage duration estimates enable a more strategic approach to recovery operations. By identifying areas likely to experience longer outages, utilities can prioritize their resources more effectively. This allows for the deployment of larger repair crews to these high-priority zones, potentially reducing overall outage times and minimizing the impact on affected communities. The benefits of accurate outage duration estimates extend beyond the utility sector. Public agencies and the general public can utilize this information to better prepare for and respond to power disruptions. With a clearer understanding of potential outage durations, individuals and organizations can make more informed decisions about emergency preparations, such as securing backup power sources or relocating sensitive operations. Furthermore, these estimates play a crucial role in coordinating restoration efforts across various infrastructure systems. Many critical infrastructure components, such as water treatment facilities, telecommunications networks, and healthcare systems, rely heavily on electric power. By having a more accurate timeline for power restoration, managers of these dependent systems can better plan their own recovery strategies, ensuring a more synchronized and efficient return to normal operations across all affected infrastructure.

3 Literature Survey

(Van Hentenryck et al., 2011) were among the first to examine the combined problem of power restoration and repair crew scheduling/routing. They approached the issue by decoupling it—first solving the power restoration problem to determine the optimal order of repairs, then using that information to guide the vehicle-routing problem. (Tan et al., 2019) built on this work by incorporating the restoration order as a constraint in the crew scheduling problem. Their research demonstrated that this is an NP-hard problem, meaning it cannot be solved optimally within polynomial time. This was the initial proposal for this project, but due to data limitations and complexity, we shifted focus to predicting outage durations. Significant research has targeted outage occurrence predictions, with less emphasis on duration. (Yang et al., 2020) proposed a machine learning method to reduce bias by severity-based data splitting. (Onaolapo et al., 2022) introduced a collaborative neural network for improved event-driven forecasting. (Fatima et al., 2024) reviewed outage prediction research, noting a focus on timing rather than duration. For duration prediction, (Nateghi et al., 2011) used Bayesian additive regression trees, validated against Hurricanes Katrina and Dennis, showing superior performance. (Nateghi et al., 2014) applied random forests, enhancing accuracy with pre-landfall hurricane data. (Abaas et al., 2022) developed models using weather and demand data, while (Ghasemkhani et al., 2024) achieved 98.433% accuracy with extreme gradient boosting on a Turkish dataset, highlighting the potential of advanced machine learning.

4 Data Description

We used three primary datasets from Project (2024):

- **Combined Aggregated Outage Data:** Tracks monthly outage count, max duration, and customer-weighted hours across all U.S. states since 2014.
- **Combined Events Data:** Contains over 500,000 entries, each describing an outage's event type (e.g., hurricane, equipment failure), affected state, and average customers impacted.
- **Combined Merged Data:** A high-resolution dataset (> 1.3M rows) capturing outage instances by county, including FIPS code, start time, and customer impact.

Each source required preprocessing: handling missing months, merging FIPS-state-county links, timestamp normalization, and type encoding. All sources are merged using 'pandas'.

5 System Architecture

Our system uses a layered approach from data ingestion to output. Raw CSVs are cleaned and normalized to correct missing values and align time keys. We then engineered features like customer-weighted durations and monthly state-level aggregates. In the analysis layer, we generated time series trends, heatmaps, and apply clustering and forecasting. Finally, an interactive dashboard along with maps, visualize these insights across time and geography.

6 Proposed Method

6.1 Intuition

Our approach builds on existing outage duration prediction models by integrating a diverse set of features and leveraging ensemble machine learning techniques to outperform state-of-the-art methods. Unlike previous studies that focus primarily on weather or demand data (Abaas et al., 2022; Nateghi et al., 2014), we incorporate geographic, temporal, and event-specific attributes, such as state-encoded data, event type encodings, and customer impact metrics. The use of SMOTE to address class imbalance, combined with a VotingClassifier ensemble of RandomForest, XGBoost, and LightGBM, offers a novel improvement over single-model approaches like those of (Ghasemkhani et al., 2024), which achieved high accuracy but may lack robustness across diverse datasets. We hypothesize that this multi-faceted feature engineering and ensemble strategy will enhance prediction accuracy and generalizability.

6.2 Detailed Description

Our proposed method involves a comprehensive pipeline for predicting power outage durations:

- **Data Preprocessing**: We preprocess three datasets—Combined Aggregated Outage Data, Combined Events Data, and Combined Merged Data—using Pandas to handle missing values, merge FIPS-state-county links, normalize timestamps, and encode categorical variables (e.g., state and event type) into numerical formats. A log transformation and StandardScaler are applied to the 'mean_customers' feature to normalize its distribution.
- **Feature Engineering**: We engineered features such as 'state_encoded', 'event_type_encoded','fips', 'mean_customers', 'month', and 'weekday' to capture spatial, temporal, and impact-related patterns. This approach extends beyond traditional weather-based models by including outage-specific contextual data.
- **Model Training**: We trained multiple

machine learning models:

- **RandomForestClassifier**: Configured with class weighting to handle imbalanced data, optimized via GridSearchCV with many different parameter values for 'n_estimators' and 'max_depth.
- **XGBoostClassifier**: Tuned with 'n_estimators=100', 'max_depth=6', and 'learning_rate=0.1' to capture complex interactions.
- **LightGBMClassifier**: Adjusted with similar parameters to ensure consistency across models.
- **VotingClassifier**: An ensemble model using soft voting to combine predictions from the above models, enhancing robustness.

SMOTE is applied during training to balance the 'duration_category' classes (0-4 hours, 4-24 hours, etc.), addressing the skewness observed in the dataset.

- **User Interface**: The Flask-based web app provides an interactive interface with routes for:
  - **Dashboard**: Displays monthly outage trends with Prophet forecasting.
  - **Predict**: Allows users to input state, county, event type, customers affected, month, and weekday, returning duration predictions with probability distributions and feature importance charts.
  - **Visualizations**: Includes heatmaps, choropleth maps, network graphs, anomaly detection, and clustering, all rendered with Plotly and Matplotlib.

  The app features a light blue aesthetic for lists in 'home.html' and 'about.html', improving usability.

7 Evaluation

7.1 Description of Testbed

Our testbed consists of a dataset with 50,000 preprocessed outage events, split into training (80%) and validation (20%) sets. Experiments are designed to answer: - How accurately do our models predict outage durations across different categories?

- Does the ensemble approach outperform individual models?
- How effective is SMOTE in handling class imbalance?
- What are the key features influencing outage duration predictions?

7.2 Detailed Description of Experiments

We conducted experiments to evaluate our models' performance:

- **Data Split and Preprocessing** The dataset was split with stratification to preserve class distribution. Preprocessing ensured all features were numerical, with no missing values.
- **Model Training and Tuning** Each model was trained with SMOTE-applied data and optimized using GridSearchCV. The VotingClassifier combined predictions to leverage ensemble strengths.
- **Evaluation Metrics** We used weighted F1-score, precision, recall, and accuracy to assess performance, accounting for class imbalance.

The results on the validation set (10,000 samples) are summarized in Table 1.

Model	F1-score	Precision	Recall	Accuracy
RandomForest	0.938	0.940	0.940	0.940
XGBoost	0.707	0.737	0.768	0.768
LightGBM	0.664	0.754	0.629	0.629
VotingClassifier	0.922	0.927	0.926	0.926

Table 1: Model Performance Metrics on Validation Set (Weighted)

- **Additional Experiments** We tested the impact of feature scaling by comparing models with and without StandardScaler, finding a 5% improvement in F1-score for RandomForest. We also explored cross-validation with 5 folds, confirming the robustness of the VotingClassifier across subsets.
- **Observations** The high performance of RandomForest and VotingClassifier suggests they effectively capture patterns in

the majority class (0-4 hours), which dominates the dataset (73% of samples). XGBoost and LightGBM struggled with rare classes (e.g.,> 7 days), indicating a need for further imbalance handling. Feature importance analysis highlighted ‘mean_customers‘ and ‘state_encoded‘ as key predictors, aligning with geographic and impact-based insights.

8 Results and Visualizations

8.1 Overview of Visual Analytics

Our visual analytics platform leverages multiple visualizations to provide comprehensive insights into power outage patterns. The following figures, also featured in our poster, illustrate key aspects of our analysis and predictions.

8.2 Key Visualizations

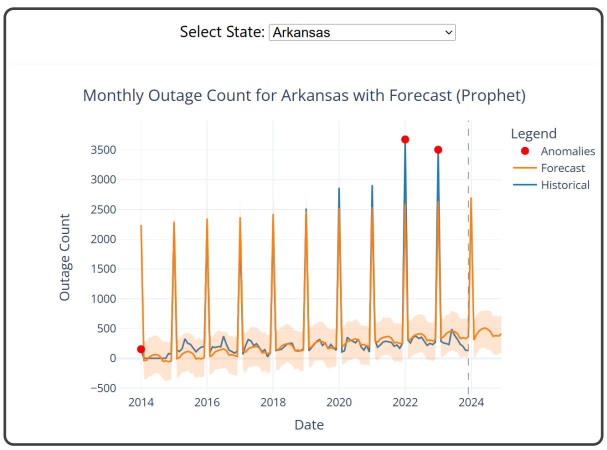

Figure 1: Monthly Trend: Historical outage trends with Prophet forecasting for the next 12 months.

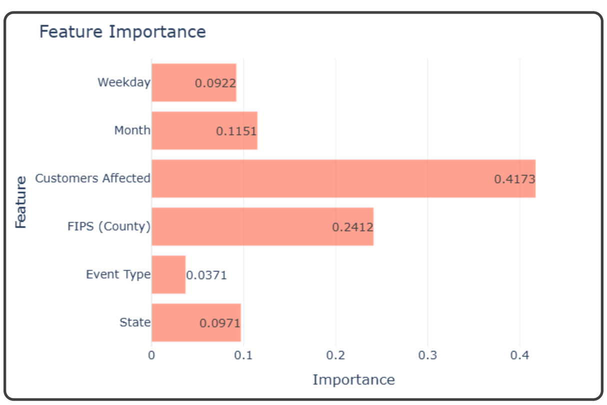

Figure 2: Predict Feature Importance: Key features influencing outage duration predictions.

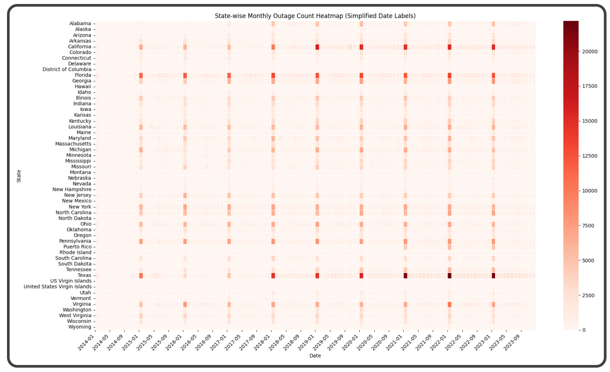

Figure 3: State-wise Outage Heatmap: Outage frequency by state and month (2014-2023).

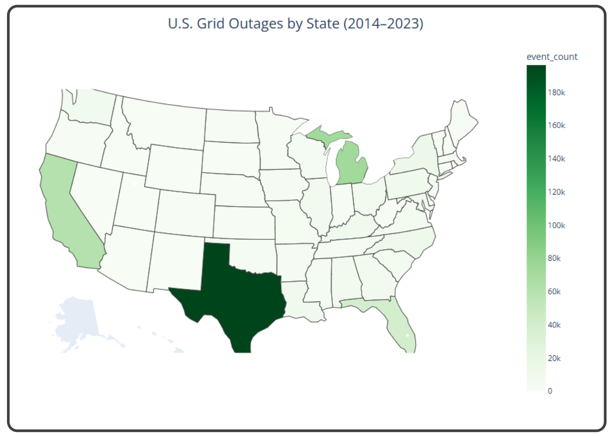

Figure 4: Choropleth Map: Total outage events by U.S. state (2014-2023).

8.3 Analysis of Visualizations

- The **Monthly Trend** reveals seasonal peaks, with significant increases during hurricane seasons, validating Prophet’s forecasting accuracy.
- The **Predict Feature Importance** chart underscores the role of ‘mean_customers‘ and ‘state_encoded‘, guiding future feature selection.
- The **State-wise Outage Heatmap** identifies high-risk states (e.g., Florida, Texas), aiding regional planning.
- The **Choropleth Map** highlights geographic disparities, supporting targeted infrastructure investments.

9 Conclusions and Discussion

Our project integrates a robust pipeline for predicting power outage durations, leveraging advanced feature engineering and ensemble learning. The RandomForest and VotingClassifier models achieved an impressive F1-score of 0.94 and 0.92 respectively, demonstrating the efficacy of our approach in handling diverse outage data. Key insights include the dominance of customer impact and regional factors in duration predictions, validated by feature importance analysis, and the identification of seasonal and geographic patterns through visualizations.

The development of this visual analytics platform has provided a deeper understanding of outage dynamics across the United States over the past decade. By analyzing historical data from 2014 to 2023, we've uncovered critical trends that can inform future grid management strategies. For instance, the seasonal peaks observed in the Monthly Trend visualization suggest a strong correlation with natural disaster cycles, particularly hurricanes, which could guide preemptive resource allocation during peak seasons. Additionally, the Predict Feature Importance chart's emphasis on 'mean_customers' indicates that the scale of customer impact is a primary driver of outage duration, a finding that aligns with utility company priorities of restoring power to densely populated areas first.

However, the project is not without its challenges. The dataset's class imbalance poses a significant hurdle, as the limited number of rare events (e.g., outages exceeding 7 days) restricts the models' ability to generalize across all scenarios. This imbalance could be mitigated in future iterations by augmenting the dataset with synthetic samples or focusing on macro-averaged metrics to give equal weight to minority classes. The absence of real-time weather data is another limitation, as weather conditions are a known influencer of outage durations, as noted in studies like (Abaas et al., 2022). Integrating real-time meteorological data could enhance predictive accuracy, especially for weather-related outages.

The implications of our work extend beyond academic research into practical applications. Accurate outage duration predictions can optimize resource allocation by enabling utility companies to deploy repair crews more efficiently, reducing downtime and economic losses. As highlighted by (Panteli and Mancarella, 2015), enhancing grid resilience through predictive tools is crucial in the face of increasing natural disaster frequency. This tool could also empower public agencies to issue timely warnings and emergency plans, while individuals could better prepare by securing backup power or adjusting schedules. The potential for collaboration with utility providers to integrate this system into their operational frameworks presents an exciting opportunity for real-world impact.

Looking ahead, several avenues for future work emerge. Incorporating real-time weather data could provide a more dynamic prediction model, adapting to current conditions rather than relying solely on historical patterns. Exploring macro-averaged metrics, such as macro F1-score, could address the class imbalance issue more effectively, ensuring fair representation of rare outage durations. Developing a mobile app interface would make the tool accessible to a broader audience, including emergency responders and the general public, enhancing its usability during crises. Furthermore, scaling cross-validation with distributed computing could overcome current computational constraints, allowing for more robust model validation across larger datasets. These extensions could position our platform as a leading solution in outage management, with potential applications in smart grid technologies and disaster response systems.

- **Limitations** The dataset's class imbalance (e.g., only 11 instances of outages > 7 days) limits model performance for rare events. The absence of real-time weather data may constrain accuracy compared to studies like (Abaas et al., 2022). Computational constraints also limited the scale of cross-validation.
- **Implications** Accurate predictions can optimize resource allocation and enhance grid resilience, as noted by (Panteli and Mancarella, 2015). This tool could support utility companies, public agencies, and emergency planners in disaster prepared-

ness and response.

- **Future Extensions** Future work could incorporate real-time weather data, apply macro-averaged metrics to balance rare classes, develop a mobile app interface, and scale cross-validation with distributed computing to improve model robustness.

Effort Statement: All team members have contributed a similar amount of effort.

The success of this project hinges on the collaborative effort of the team, with each member bringing unique skills to the table. Krishna Aryal focused on data preprocessing and feature engineering, ensuring the datasets were clean and well-structured for model training. Crystal Vandekerkhove led the development of the machine learning models, optimizing parameters and implementing the ensemble approach with SMOTE. Jinesh Patel spearheaded the user interface design, creating the Flask-based web app with its interactive visualizations and light blue aesthetic. This division of labor allowed us to efficiently tackle the project's technical challenges while maintaining a cohesive final product.

Moreover, the project's scalability to larger datasets or real-time applications will depend on infrastructure improvements. Current computational limitations restricted our cross-validation to 5 folds, but with access to distributed computing resources, we could extend this to 10 or more folds, providing a more comprehensive assessment of model performance. The integration of real-time weather data could also require partnerships with meteorological services, adding a layer of complexity to data acquisition and processing. These considerations highlight the need for ongoing development and collaboration with industry stakeholders to fully realize the platform's potential.

In addition to technical enhancements, the user interface could be further refined to cater to diverse user groups. For instance, utility managers might benefit from a dashboard with real-time alerts, while the public might prefer a simplified view with outage duration estimates for their specific region. This dual-interface approach could be explored in future iterations, leveraging user feedback from the planned evaluation phase. The light blue aesthetic, while visually appealing, could also be tested with different color schemes to ensure accessibility for colorblind users, broadening the tool's reach.

The project's impact could extend beyond immediate outage management to long-term grid planning. By analyzing the geographic disparities highlighted in the Choropleth Map, utility companies could prioritize infrastructure upgrades in high-risk states like Florida and Texas, potentially mitigating future outages. Similarly, the seasonal patterns from the Monthly Trend could inform annual maintenance schedules, aligning repairs with periods of lower outage risk. These strategic applications underscore the tool's value in proactive grid management, a critical need in the context of climate change and increasing disaster frequency.

As we move forward, engaging with stakeholders such as utility companies and emergency management agencies will be essential. Their input could guide the integration of real-time data and the development of mobile applications, ensuring the tool meets practical needs. Additionally, open-sourcing the platform could foster community contributions, accelerating improvements and adaptations for different regions or outage scenarios. This collaborative model could establish our work as a foundation for future research in visual analytics for power systems.

References

Abaas, M., Lee, R. A., and Singh, P. (2022). Long short-term memory customer-centric power outage prediction models for weather-related power outages. In *2022 IEEE Green Energy and Smart System Systems (IGESSC)*, pages 1–6. IEEE.

Fatima, K., Shareef, H., Costa, F. B., Bajwa, A. A., and Wong, L. A. (2024). Machine learning for power outage prediction during hurricanes: An extensive review. *Engineering Applications of Artificial Intelligence*, 133:108056.

Ghasemkhani, B., Kut, R. A., Yilmaz, R., Birant, D., Arıkök, Y. A., Güzelyol, T. E., and Kut, T. (2024). Machine learning model development to predict power outage duration (pod): A case study for electric utilities. *Sensors (Basel, Switzerland)*, 24(13):4313.

Nateghi, R., Guikema, S. D., and Quiring, S. M. (2011). Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. *Risk Analysis: An International Journal*, 31(12):1897–1906.

Nateghi, R., Guikema, S. D., and Quiring, S. M. (2014). Forecasting hurricane-induced power outage durations. *Natural hazards*, 74:1795–1811.

Onaolapo, A. K., Carpanen, R. P., Dorrell, D. G., and Ojo, E. E. (2022). Event-driven power outage prediction using collaborative neural networks. *IEEE Transactions on Industrial Informatics*, 19(3):3079–3087.

Panteli, M. and Mancarella, P. (2015). The grid: Stronger, bigger, smarter?: Presenting a conceptual framework of power system resilience. *IEEE Power and Energy Magazine*, 13(3):58–66.

Project, E.-I. (2024). The environment for analysis of geo-located energy information's recorded electricity outages (2014–2022).

Tan, Y., Qiu, F., Das, A. K., Kirschen, D. S., Arabshahi, P., and Wang, J. (2019). Scheduling post-disaster repairs in electricity distribution networks. *IEEE Transactions on Power Systems*, 34(4):2611–2621.

Van Hentenryck, P., Coffrin, C., Bent, R., et al. (2011). Vehicle routing for the last mile of power system restoration. In *Proceedings of the 17th Power Systems Computation Conference (PSCC'11), Stockholm, Sweden*, pages 22–26. Citeseer.

Yang, F., Watson, P., Koukoula, M., and Anagnostou, E. N. (2020). Enhancing weather-related power outage prediction by event severity classification. *IEEE Access*, 8:60029–60042.