Grid Disruption Analysis: Visualizing U.S. Grid Outages (2014–2023) and Outage Prediction using Advanced Machine Learning

Krishna Aryal Crystal Vandekerkhove Jinesh Patel
Team 155 April 2025

Motivation

Power outages caused by natural disasters impact thousands annually, posing significant challenges for utilities in restoration planning and crew coordination. This project aims to tackle these issues by developing a visual analytics platform that leverages machine learning to predict outage durations, offering critical insights to support faster, data-driven recovery efforts.

Project Overview

This project analyzes U.S. grid outages from 2014 to 2023, using interactive visualizations and machine learning to uncover patterns, predict outage durations, detect anomalies, and cluster states by outage patterns. The Flask-based web app features a user-friendly homepage with a light blue aesthetic for the list of visualizations, and a similarly styled About page detailing the machine learning models used. Key approaches include:

- **Forecasting** with Prophet to predict outage trends.
- **Anomaly Detection** using Isolation Forest to identify unusual events.
- **Clustering** with K-Means to group states by outage patterns.
- **Prediction** of outage durations using RandomForest, XGBoost, LightGBM, and VotingClassifier with SMOTE.

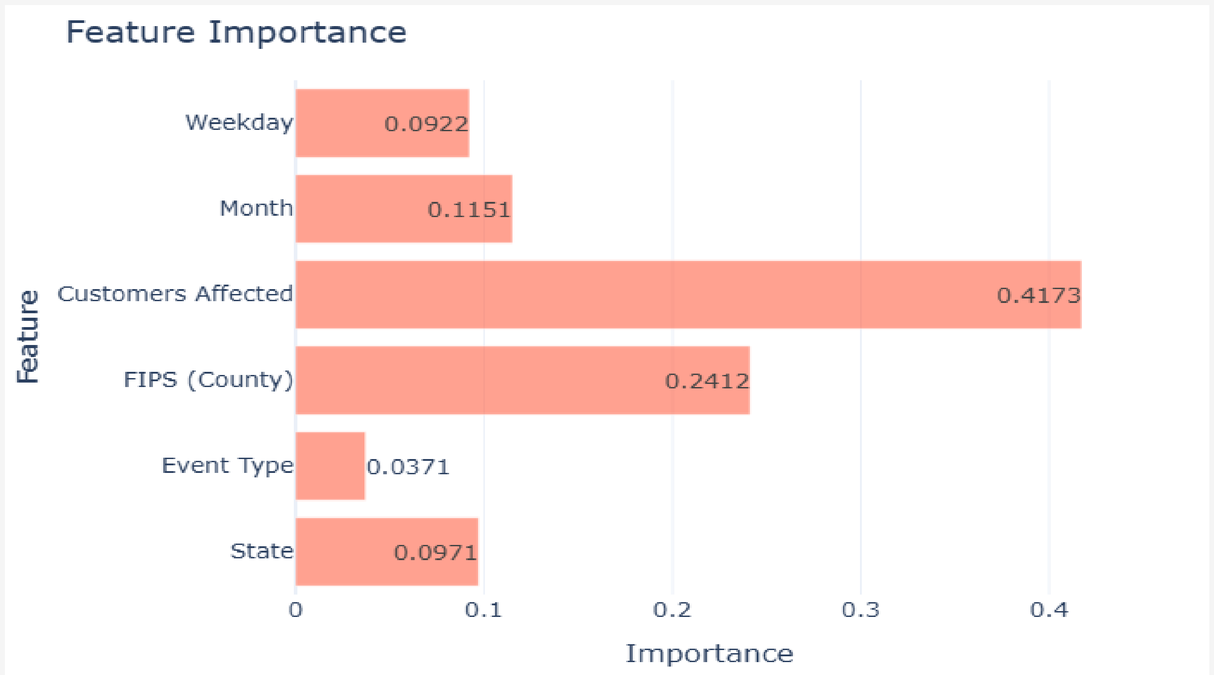
Machine Learning Models

We employ the following models to analyze and predict grid disruptions:

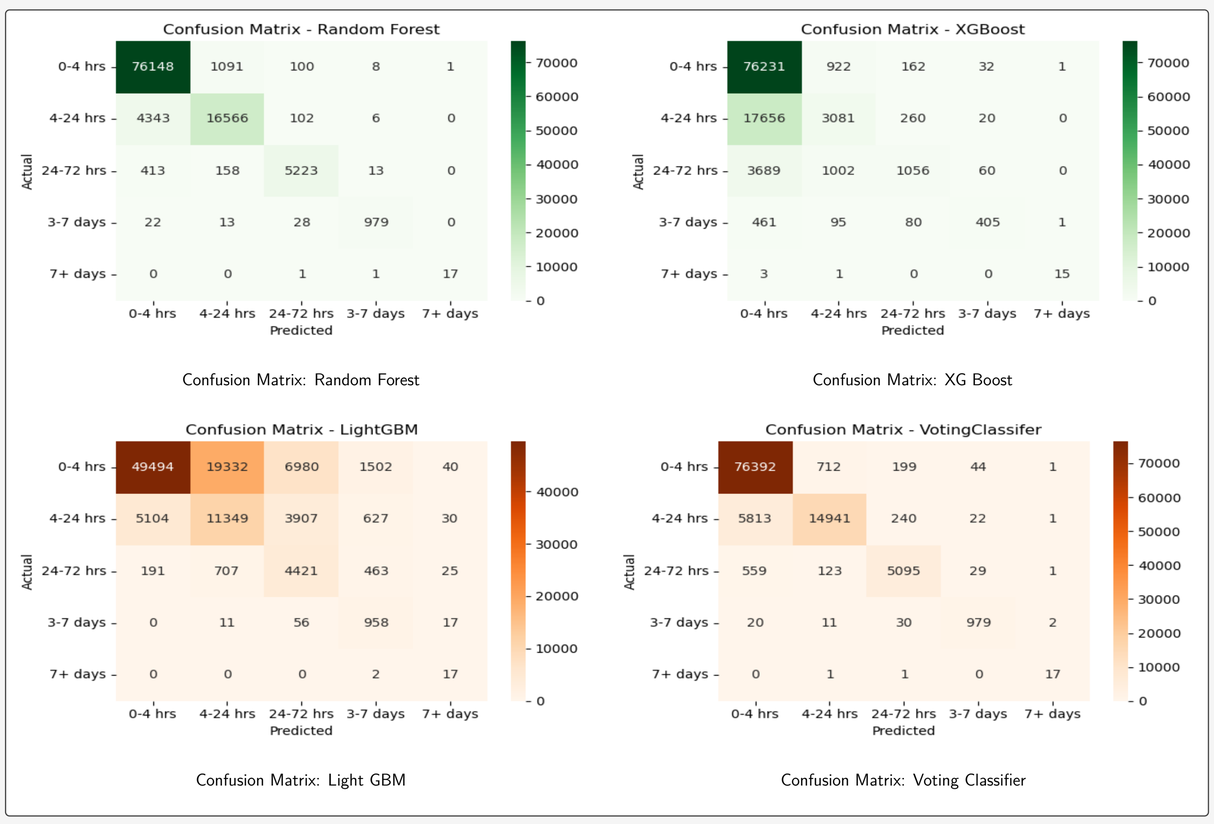
- **Prophet:** Forecasts monthly outage counts, capturing seasonality and trends.
- **Isolation Forest:** Detects anomalies in outage events, highlighting significant disruptions.
- **K-Means Clustering:** Groups states with similar outage patterns, revealing regional trends.
- **RandomForest, XGBoost, LightGBM, VotingClassifier:** Predict outage durations, trained with SMOTE to handle class imbalance.

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: Performance metrics (weighted) for various classifiers



Feature importance chart for outage duration prediction.



Confusion Matrix: Random Forest

Confusion Matrix: XG Boost

Confusion Matrix: Light GBM

Confusion Matrix: Voting Classifier

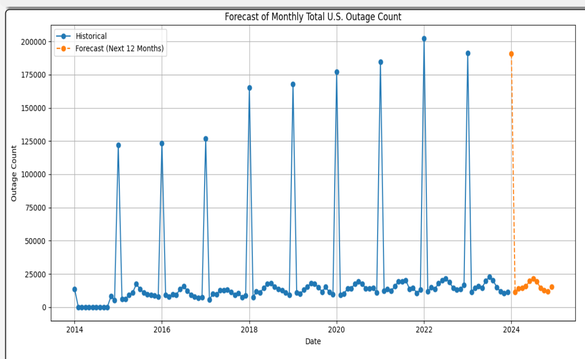
Dataset Description

The dataset comprises U.S. grid outage events from 2014 to 2023, sourced from publicly available records. It includes:

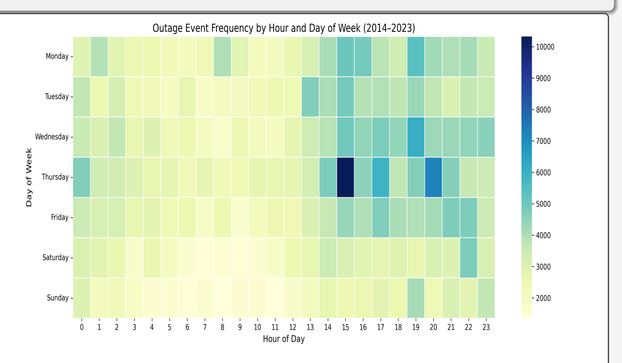
- **Time Series Data:** Monthly outage counts by state.
- **Event Data:** Detailed outage events with timestamps, states, and counties.
- **Attributes:** Outage duration, affected customers, and event types.

Preprocessing involved handling missing months, merging FIPS-state-county links, timestamp normalization, and type encoding using Pandas.

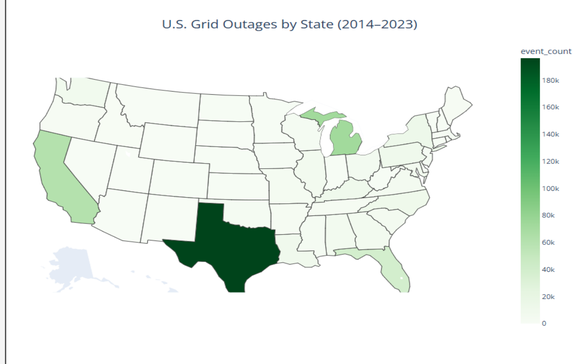
Data Visualization



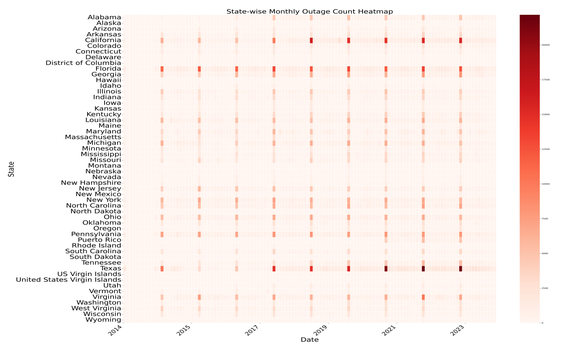
Monthly Trend: Historical outage trends with forecasting using Prophet.



Hourly Heatmap: Frequency of outages by hour of day.



Choropleth Map: Total outage events by state from 2014 to 2023.

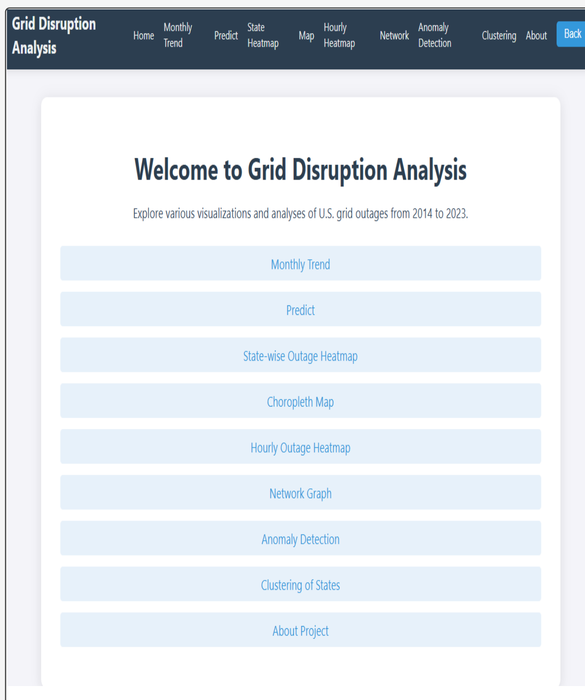


State-wise Outage Heatmap: Outages by state and month from 2014 to 2023

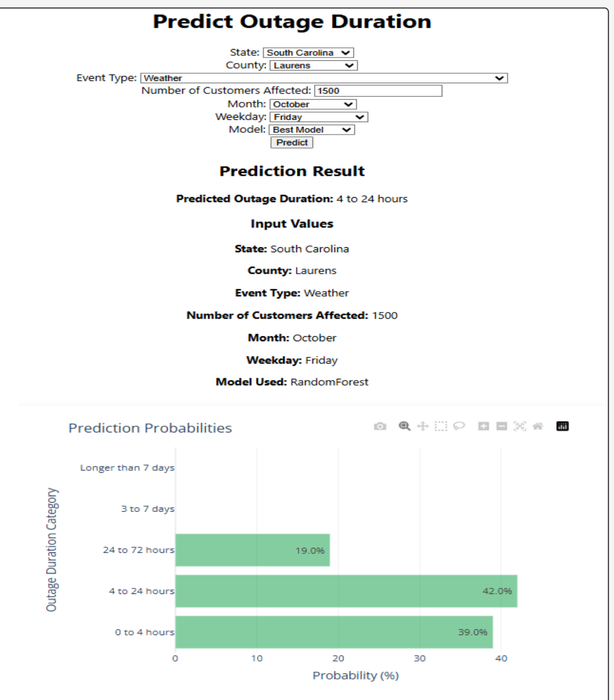
Key Visualizations

- **Monthly Trend:** Interactive trends with forecasting.
- **Predict:** Outage duration prediction with probability and feature importance charts.
- **State-wise Heatmap:** Outage counts by state and month.
- **Choropleth Map:** Visualizes outages by state.
- **Hourly Heatmap:** Highlights peak outage times.
- **Network Graph:** Shows connections between outage events.
- **Anomaly Detection:** Identifies unusual events.
- **Clustering of States:** Groups states by outage patterns.
- **About Project:** Details the machine learning models used.

User Interface



User Interface: Main Page



User Interface: Outage Prediction

Results and Impact

Our analysis provides actionable insights for grid management:

- Outages peak during certain hours and days, aiding in targeted mitigation.
- States with similar outage patterns can share resources and strategies.
- Anomalies highlight critical events for investigation.
- Accurate duration predictions (best F1-score: 0.938) improve resource allocation.
- Forecasts help anticipate future disruptions, enhancing grid resilience.