Forecasting Future Weather Impacts on Power Outages

MLPR PROJECT

Power Outage: a concern

Power outages have severe economic impacts on industries such as:

- Manufacturing
- Banking & IT services
- Critical Infrastructure (Hospitals, Fire Stations)
- Public transport (flight cancellations, train delays)

Weather-Related Power Outages

- Weather-related events account for 80% of major U.S. outages (2000–2023)
- Major hurricanes cause mass blackouts and billions in damages
- Superstorm Sandy (2012) → 8.1 million homes lost power
- Hurricane Michael (2018) → 1.7 million customers affected
- As climate change intensifies, hurricanes, heat waves, and winter storms will increase outage risks

Nat. Hazards Earth Syst. Sci., 23, 1665–1683, 2023 https://doi.org/10.5194/nhess-23-1665-2023

PROBLEM STATEMENT

Power outages caused by extreme weather can lead to significant economic losses and risks to human safety. This project aims to create a system that predicts power outages based on weather conditions. By doing so, it can help improve disaster preparedness and response, reducing the impact of these outages on people and businesses.

Literature Review

Existing Models & Approaches

Generalized Additive Model (GAM):

- Flexible regression model allowing non-linear relationships
- Power outage data during past hurricanes in the Gulf Coast region
- The explanatory variables were:
    - winds experienced
    - long-term precipitation and soil moisture levels
    - power system components
    - land use and land cover

""Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models"
– Seung-Ryong Han.

Existing Models & Approaches

Random Forest (RF):

- Uses combined outputs of multiple decision trees on random subsets of data and features
- Model trained on outages caused by Hurricanes Dennis, Katrina, and Ivan in a central Gulf Coast state
- Variables were number of customers, power system components (poles, switches, transformers), geographic and climatic factors, and hurricane characteristics (wind speed).

"Forecasting Hurricane-Induced Power Outage Durations" – Nateghi, R., Guikema, S. D., & Quiring, S. M.

Existing Models & Approaches

Bayesian additive regression trees (BART):

- Ensemble method that combines multiple shallow trees to model complex relationships while providing uncertainty estimates.
- power outages caused by Hurricane Ivan in 2004.
- Better prediction accuracy than accelerated failure time (AFT) and Cox proportional hazard models (Cox PH)

.

‘Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes’-Seth et al.-

GAPS IN CURRENT METHODOLOGIES

Over-Reliance on High-Impact Events:

Many models are predominantly trained on severe storms (e.g., hurricanes, blizzards, tornadoes). This focus can limit their ability to produce realistic operational outage predictions for power utilities and emergency managers, leading to over-predictions for moderate storms and reducing practical utility.

Watson, P., & Rajagopalan, S. (2022). Improved quantitative prediction of power outages caused by extreme weather events. International Journal of Forecasting, 38(2), 789-805

Why is this a problem?

Over-Predictions for Moderate Events: When models are trained mostly on severe weather, they tend to overestimate the risk and impact of less severe weather, such as moderate thunderstorms or strong winds.

Lack of Focus on Snow & Ice-Related Outages:
While most research emphasizes storm-related outages, wet snow and icing events remain underexplored despite their significant impact on power grids.

Cerrai, D. (2019). Predicting weather-caused power outages: Technique development, evaluation, applications. Doctoral Dissertation, University of Connecticut.

Why is this a problem?

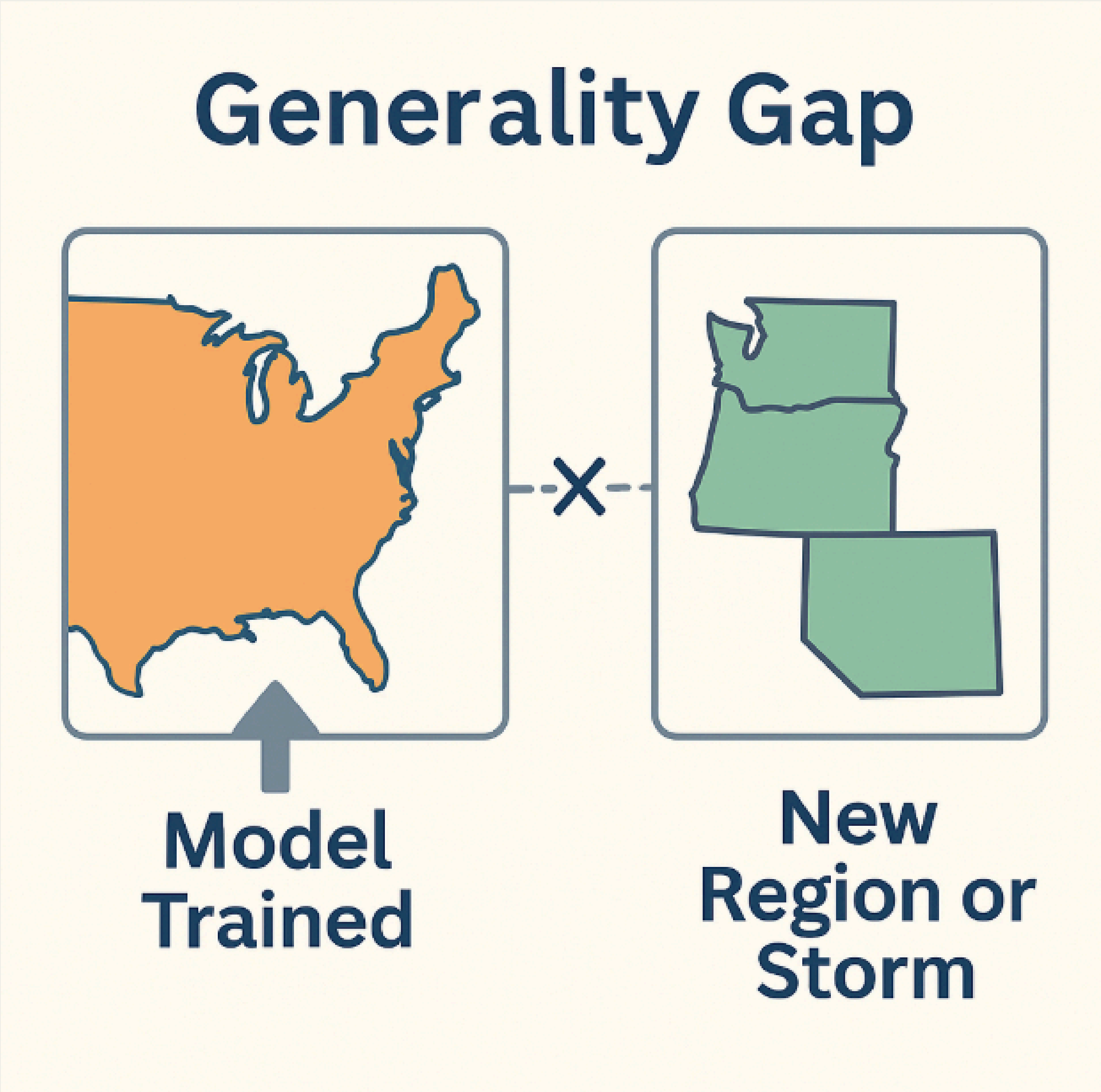
Underserved Cold-Climate Regions: Many power grids in colder regions are severely affected by freezing rain, wet snow, and ice accumulation, which can weigh down power lines and topple poles.

Generality and Transferability Gap:

Certain models, like those of Liu et al., are designed for individual utilities or storm indicators and are not transferable to other events or general outage prediction. Others, like that of Han et al., utilize generalizable features but are, nevertheless, limited by the nonlinear and location-based nature of outages.

Why is this a problem?

The current generation of ML models remains heavily event-specific, lacking the breadth and adaptability required for comprehensive outage prediction across climate scenarios.



Our Model

Data Collection

We began our project with two primary datasets:

- NOAA Storm Events Dataset
    - Source: NOAA Storm Events
        - Features included: Event_Type, Event_Date, Event_Time, Event_Latitude, Event_Longitude, Event_State, Event_County

- Power Outage Dataset (Eagle-I)
    - Source: OSTI Power Outage Dataset
        - Features included: Outage_Date, Outage_Time, Outage_State, Outage_County

While the storm event data gave us the location and time of weather incidents, it lacked detailed atmospheric or environmental variables (e.g., windspeed, humidity, temperature) that are often important predictors in outage modeling. This limitation made it insufficient for training a machine learning model.
We had data in sentences that contains meaningful information about the severity of the event but we don’t know how to extract the data in a way that can be interpreted by our ML model.

Weather Data Enrichment Using Open-Meteo API

To enrich the storm event data with weather variables, we used the Open-Meteo Historical Weather API. This API allows us to extract hourly weather conditions based on a specific time and location.

Initial Limitation and API Key Upgrade

- Initially, we had access to only 15,000 data requests through the public Open-Meteo API. This constraint limited the scale and coverage of our enrichment process.
- To overcome this, we contacted the Open-Meteo team directly and were generously provided with a free API key that allowed unlimited access to historical weather data. This upgrade enabled us to fully enrich the entire dataset without any restrictions or cost.

Data Extract Process Using Open-Weather API

Had:

- Event_Type
- Event_Date
- Event_Begin_Time
- Event_End_Time
- Event_Latitude
- Event_Longitude
- Event_State
- Event_County

Required:

API Key: xWPlcPnPJKtpKga2

Literature:

Chen, H., Zhang, Y., Wu, Y., & Schofield, R. J. R. (2020).
Dynamic Modeling of Power Outages Caused by Thunderstorms.
Informatics, 2(2), 151–168.
https://doi.org/10.3390/informatics2020010

Heiskanen, L., Rosti, J., Heikkinen, T., et al. (2019).
Short-term Prediction of Electricity Outages Caused by Convective Storms.
arXiv preprint arXiv:1907.00662
https://arxiv.org/abs/1907.00662

Kurukuru, K., Bhattacharya, S., Baran, M., & Chakrabortty, A. (2022).
Forecasting Weather-Related Power Outages Using Weighted Logistic Regression.
IET Smart Grid, 5(1), 38–46.
https://doi.org/10.1049/stg2.12109

Henriet, B. C., Schnell, J. C., Pegion, P., et al. (2020).
The Occurrence and Variability of Major Electrical Power Disruptions in the United States.
Energy, 213, 118662.
https://doi.org/10.1016/j.energy.2020.118662

Delogu, M., Viola, F., & Burlando, P. (2023).
Hail: Mechanisms, Monitoring, Forecasting, Damages, Financial Compensation Systems, and Prevention.
Atmosphere, 14(11), 1642.
https://doi.org/10.3390/atmos14111642

Lan, T., Zhang, L., Liu, X., et al. (2022).
Thunderstorm Nowcasting with Deep Learning: A Multi-Hazard Data Fusion Model.
arXiv preprint arXiv:2211.01001
https://arxiv.org/abs/2211.01001

Lu, Y. (2022).
Predicting Weather-Caused Power Outages: Technique Development, Evaluation, Applications.
Doctoral Dissertation, University of Connecticut.
https://digitalcommons.lib.uconn.edu/dissertations/2110/

Kiaghadi, A., & Rifai, H. S. (2021).
Community-Scale Big Data Reveals Disparate Impacts of the Texas Winter Storm of 2021 and Its Managed Power Outage.
arXiv preprint arXiv:2108.06046
https://arxiv.org/abs/2108.06046

Kamarthi, R. T., & Saha, P. K. (2018).
A Multi-Variable Stacked Long-Short Term Memory Network for Wind Speed Forecasting.
arXiv preprint arXiv:1811.09735
https://arxiv.org/abs/1811.09735

Kurukuru, K., Bhattacharya, S., Baran, M., & Chakrabortty, A. (2022).
(Same as above - used also under Tornado)
https://doi.org/10.1049/stg2.12109

Variables Extracted:

- temperature_2m
- dew_point_2m
- relative_humidity_2m
- precipitation
- rain
- snowfall
- snow_depth
- windspeed_10m
- windspeed_100m
- winddirection_10m
- winddirection_100m
- windgusts_10m
- surface_pressure
- cloudcover
- cloudcover_low
- cloudcover_mid
- cloudcover_high
- soil_temperature_0_to_7cm
- soil_temperature_7_to_28cm
- soil_temperature_28_to_100cm
- soil_temperature_100_to_255cm
- soil_moisture_0_to_7cm
- soil_moisture_7_to_28cm
- soil_moisture_28_to_100cm
- soil_moisture_100_to_255cm

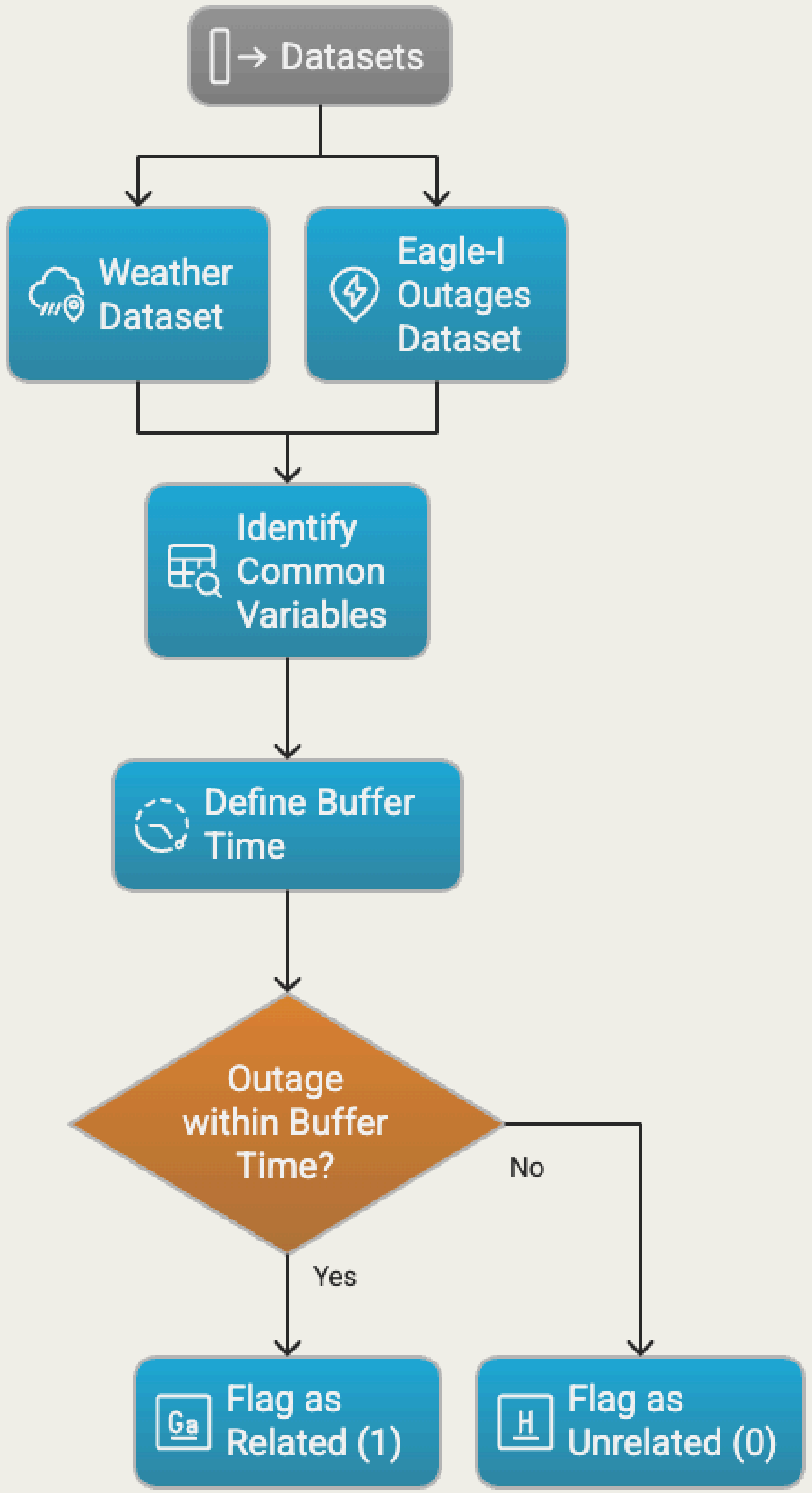
Weather Events:

- Thunderstorm
- Tornado
- High_Winds
- Heavy_Snow
- Hail

Now, we had two datasets:

- Weather Dataset – containing extracted weather features.
- Eagle-I Outages Dataset – recording instances of power outages.

- The common variables across both datasets are State, County, and Time.
- To establish a meaningful connection between weather events and power outages, we implemented the following approach: We introduced a buffer time—a defined time window starting from the occurrence of a weather event.
- Within this window, we checked whether a power outage occurred in the same state and county as the weather event. If the outage occured within the buffer time of the occurance of the weather event, we marked it as 1, else 0.



Final Dataset

tempe	dew_p	relativ	precip	rain	snowfa	snow_	windsp	windsp	winddi	winddi	windgu	surfac	cloudc	cloudc	cloudc	cloudc	soil_te	soil_te	soil_te	soil_te	soil_m	soil_m	soil_m	soil_m	event_	event_	event_	time	state	county	caused
9.7	5.4	74	0	0	0		43.6	60.3	188	190	70.2	1009.3	97	25	22	94	9.1	8.3	7.9	8.1	0.135	0.137	0.144	0.145	######	44.603	-124.0	######	nebras	dundy	0
3.2	1.6	90	0.7	0.7	0	0	35.9	55.5	312	314	64.1	889.5	97	91	82	29	8.1	14.7	11.5	5.1	0.222	0.09	0.085	0.121	######	41.024	-100.7	######	pennsy	lehigh	0
13.9	9.6	75	0.9	0.9	0	0	25.6	45.1	281	281	82.4	923.2	79	61	56	0	11.6	7.9	3.4	4.8	0.489	0.482	0.488	0.479	######	37.058	-80.72	######	ohio	darke	0
14.8	13.1	90	0	0	0	0	17.1	29.6	221	222	58	931.7	44	22	34	3	9.4	5.3	1.6	3.7	0.446	0.487	0.473	0.482	######	37.833	-79.46	######	south	hermo	0
14	12.7	92	0	0	0	0	24.2	40.8	222	223	52.2	937.7	90	89	26	0	12.1	9.6	5	6.2	0.49	0.486	0.472	0.498	######	36.681	-80.28	######	monta	fergus	0
15.1	12.9	87	0	0	0	0	22	38	215	215	48.2	974.7	100	100	91	0	13	10.8	6.6	8.5	0.481	0.481	0.48	0.42	######	36.747	-78.94	######	virginia	pulask	1
13.1	12.9	98	0.3	0.3	0	0	17.3	30.5	211	213	39.6	954.6	100	100	94	7	11.4	9.5	5.6	6.8	0.495	0.484	0.487	0.491	######	36.199	-81.134	######	texas	sherm	0
3.5	1.4	86	0	0	0	0	9.9	18.8	57	83	23.4	932.9	76	4	17	70	4.7	6.5	3.5	7.5	0.29	0.279	0.204	0.172	######	38.897	-98.79	######	north	allegha	0

No. of features: ~30

No. of:Rows:293875

Data Preprocessing

- Parsed Datetime: Extracted month, year, hour from event_datetime for temporal context.
- Scaled Features: Standardized using StandardScaler for consistency.
- Dropped Non-Numeric: Kept ~25 numeric features, excluded state, county, event_type.
- Dropped Missing Values: Removed rows with NaN values.

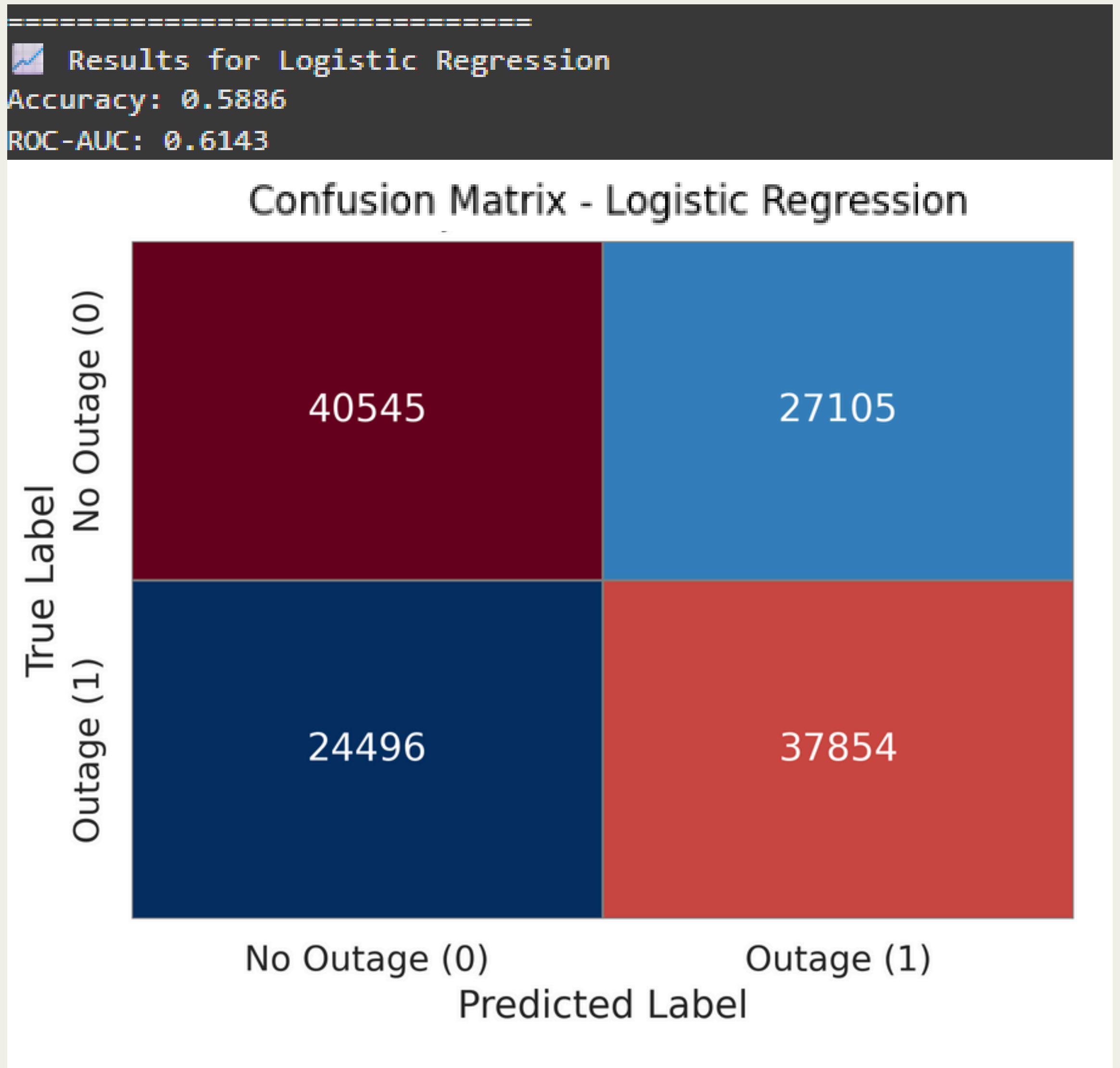
Why Avoided PCA/LDA?

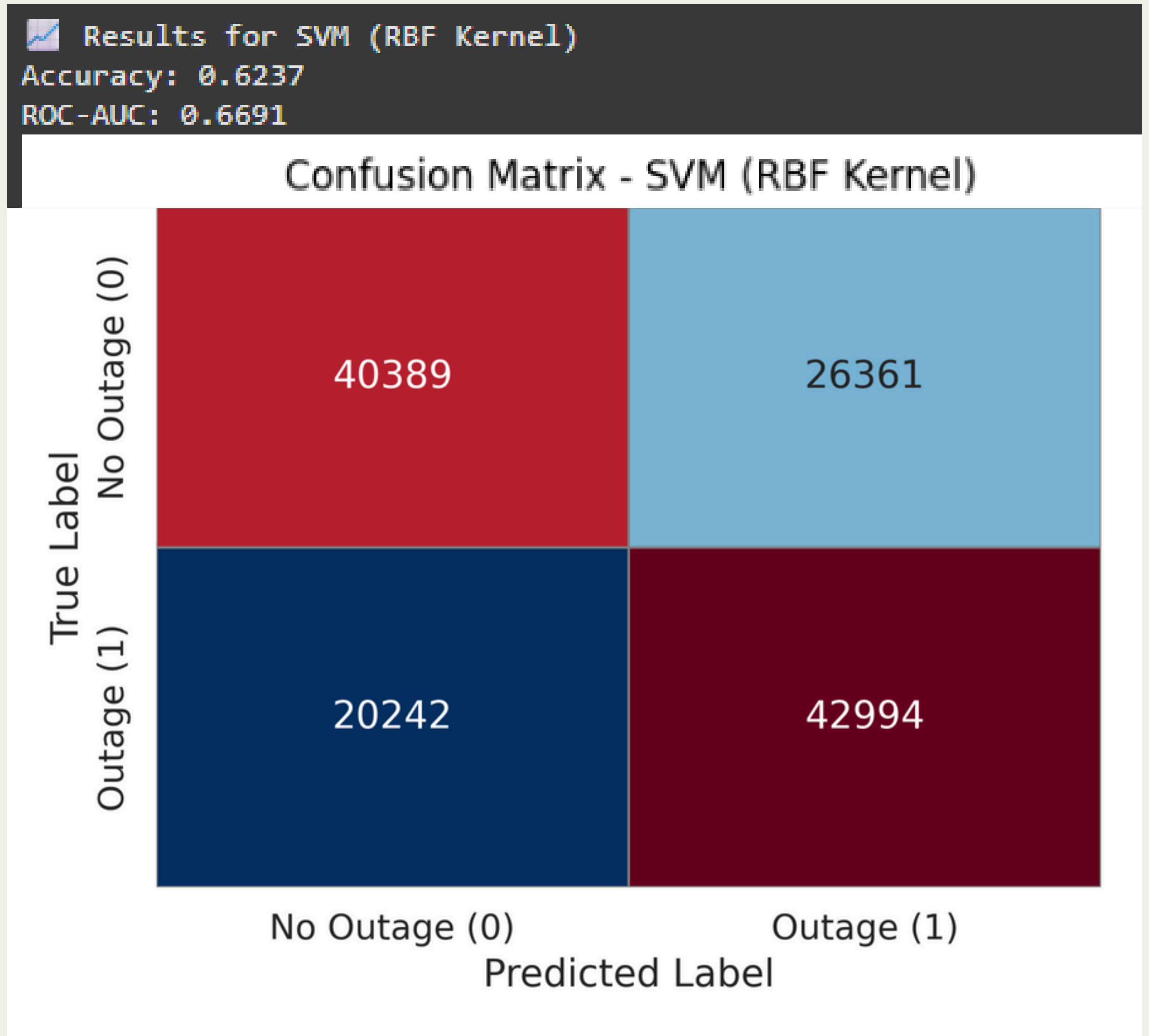
- Random Forest handles original features and selects important ones.
- PCA/LDA reduces interpretability of weather features.
- Moderate feature count (25) didn't require reduction.

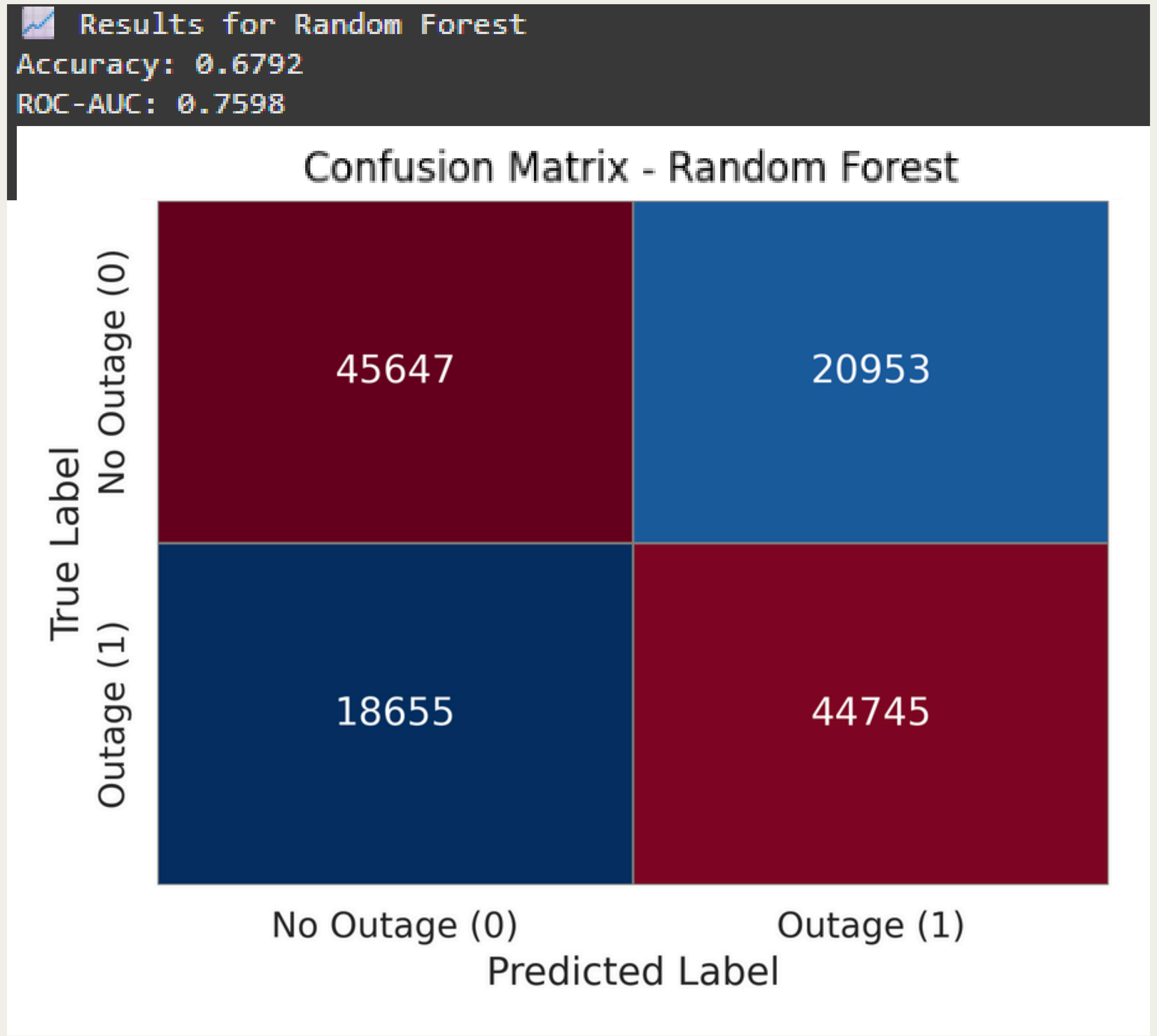
```
Missing Values:
temperature_2m            0
dew_point_2m              0
relative_humidity_2m      0
precipitation             0
rain                      0
snowfall                  0
snow_depth              689
windspeed_10m             0
windspeed_100m            0
winddirection_10m         0
winddirection_100m        0
windgusts_10m             0
surface_pressure          0
```

ML Methodology

Models Comparison







Selected Model: Random Forest (RF)
RF outperformed others in terms of AUC-ROC and accuracy
Supported by prior research on outage modeling (Wedagedara et al., 2022)

Why Random Forest?

- Ensemble of decision trees built on random subsets of data and features
- Aggregates outputs of individual trees → higher accuracy, lower variance
- Handles imbalanced, noisy, and high-dimensional datasets effectively
- Demonstrated strong AUC-ROC (≈0.75) in prior power outage studies

Models and Performance

Attempt 1: Single Model for all Weather Events

- Using our dataset we trained a random forest classifier to predict power outages without separating weather events.
- Based on literature review we used ROC-AUC as our primary metric.
- For this model we got an AUC score of 0.899

```
Top 10 Important Features:
1. event_longitude: 0.1063
2. surface_pressure: 0.0701
3. event_latitude: 0.0692
4. soil_moisture_100_to_255cm: 0.0536
5. soil_moisture_28_to_100cm: 0.0463
6. soil_temperature_100_to_255cm: 0.0457
7. dew_point_2m: 0.0447
8. soil_temperature_28_to_100cm: 0.0405
9. soil_moisture_7_to_28cm: 0.0393
10. soil_temperature_7_to_28cm: 0.0391
```

```python
# Classifier
rf = RandomForestClassifier(random_state=42, class_weight='balanced')

# Randomized search CV
random_search = RandomizedSearchCV(
    estimator=rf,
    param_distributions=param_dist,
    n_iter=30,
    scoring='roc_auc',
    cv=3,
    verbose=2,
    n_jobs=-1,
    random_state=42
)

# Fit on your scaled, numeric dataset (X_train, y_train)
random_search.fit(X_train, y_train)

# Evaluate
best_model = random_search.best_estimator_
y_pred = best_model.predict(X_test)
y_proba = best_model.predict_proba(X_test)[:, 1]

print("Classification Report:\n", classification_report(y_test, y_pred))
print("AUC-ROC Score:", roc_auc_score(y_test, y_proba))

# Confusion matrix
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm)
disp.plot(cmap="Blues")
plt.show()
```

. A hybrid machine learning approach for predicting power outages due to extreme weather events. Wedagedara, S., Perera, S., & Perera, H. N. (2022)

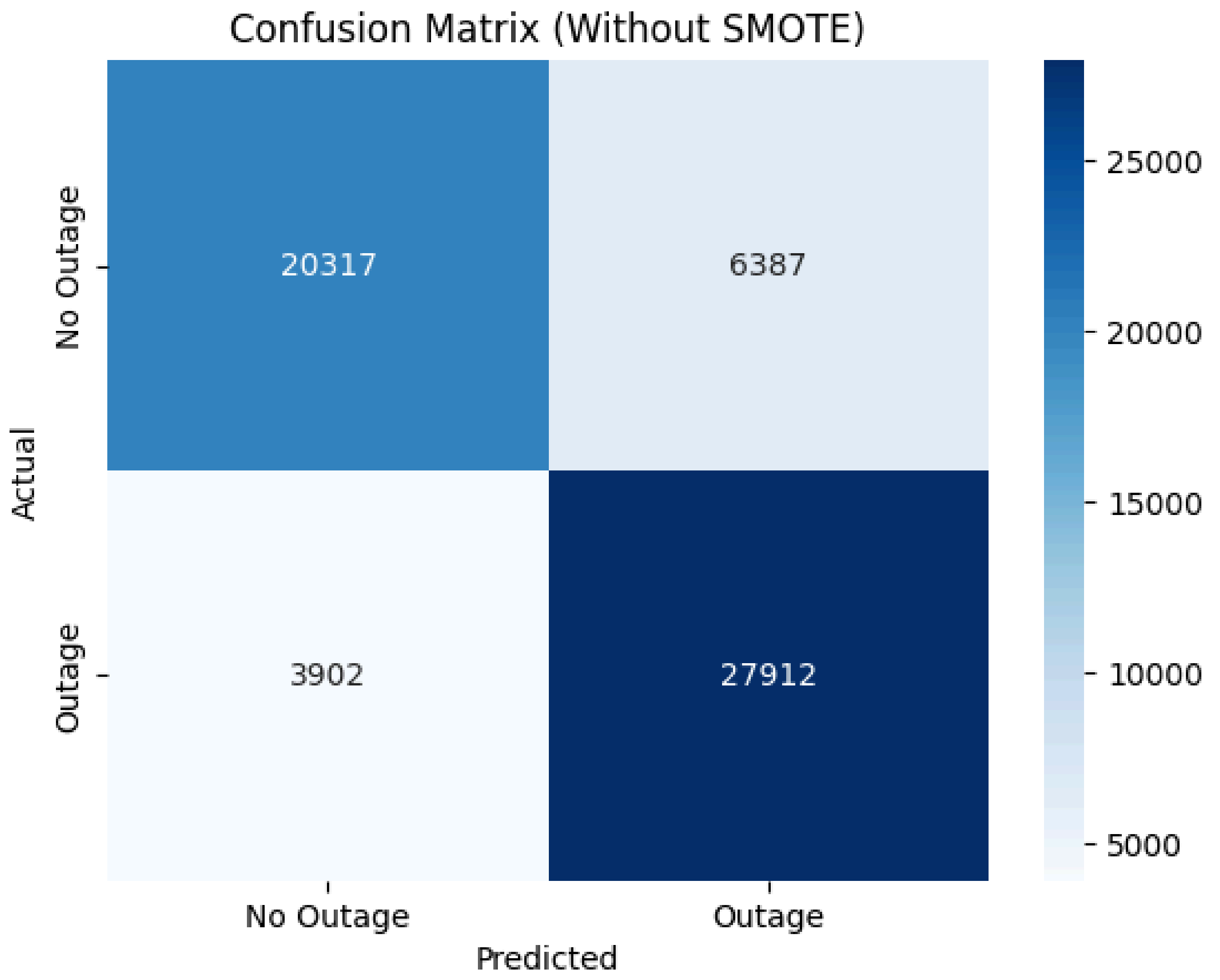
SMOTE vs No SMOTE: Impact on Model Performance

Class Distribution:
No Outage (0): 56%
Outage (1): 44%

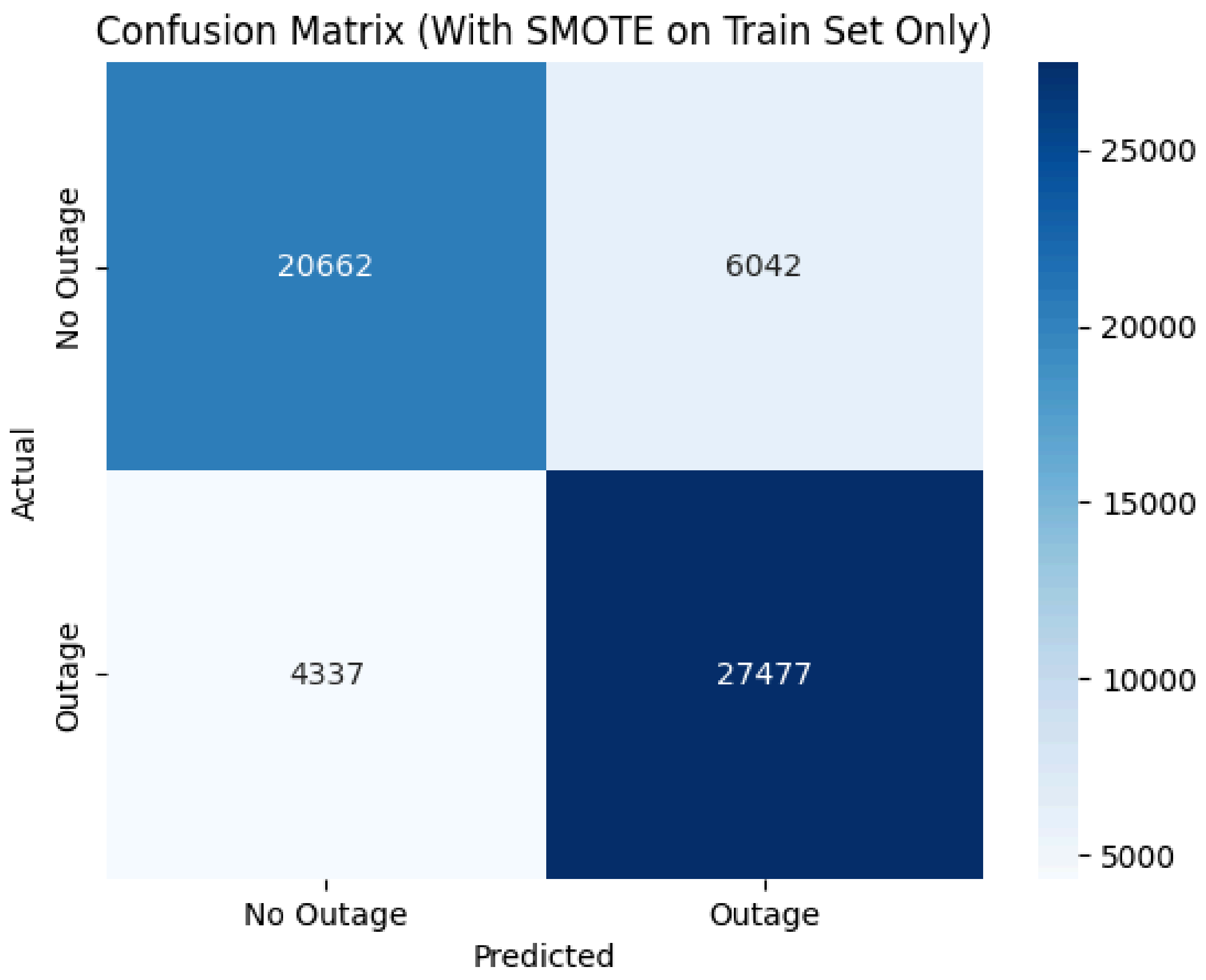
✅**Without SMOTE:**
Better accuracy, AUC, and F1-score
Handled mild imbalance well
No synthetic noise
❌ **With SMOTE:**
Slight performance drop
Synthetic data added redundancy
Less generalizable

Conclusion:
SMOTE was unnecessary — original data gave better results.

```
---- Without SMOTE ----
Confusion Matrix:
 [[20317  6387]
 [ 3902 27912]]

Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.76      0.80     26704
           1       0.81      0.88      0.84     31814

    accuracy                           0.82     58518
   macro avg       0.83      0.82      0.82     58518
weighted avg       0.83      0.82      0.82     58518

Accuracy: 0.8241737585016576
AUC-ROC Score: 0.8999226384030485
```




```
---- With SMOTE (on train set only) ----
Confusion Matrix:
 [[20662  6042]
 [ 4337 27477]]

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.77      0.80     26704
           1       0.82      0.86      0.84     31814

    accuracy                           0.82     58518
   macro avg       0.82      0.82      0.82     58518
weighted avg       0.82      0.82      0.82     58518

Accuracy: 0.8226357701903688
AUC-ROC Score: 0.8994402080972976
```



Hyperparameter Tuning of Random Forest

Objective

- Optimize the Random Forest model for better predictive performance on outage classification.
- Focused on maximizing AUC (Area Under the ROC Curve) for imbalanced data.

Method Used

Random Sampling of 30 hyperparameter combinations using ParameterSampler.
3-fold Cross-Validation with AUC as the scoring metric.
Used class_weight='balanced' to handle class imbalance.

Parameters Tuned

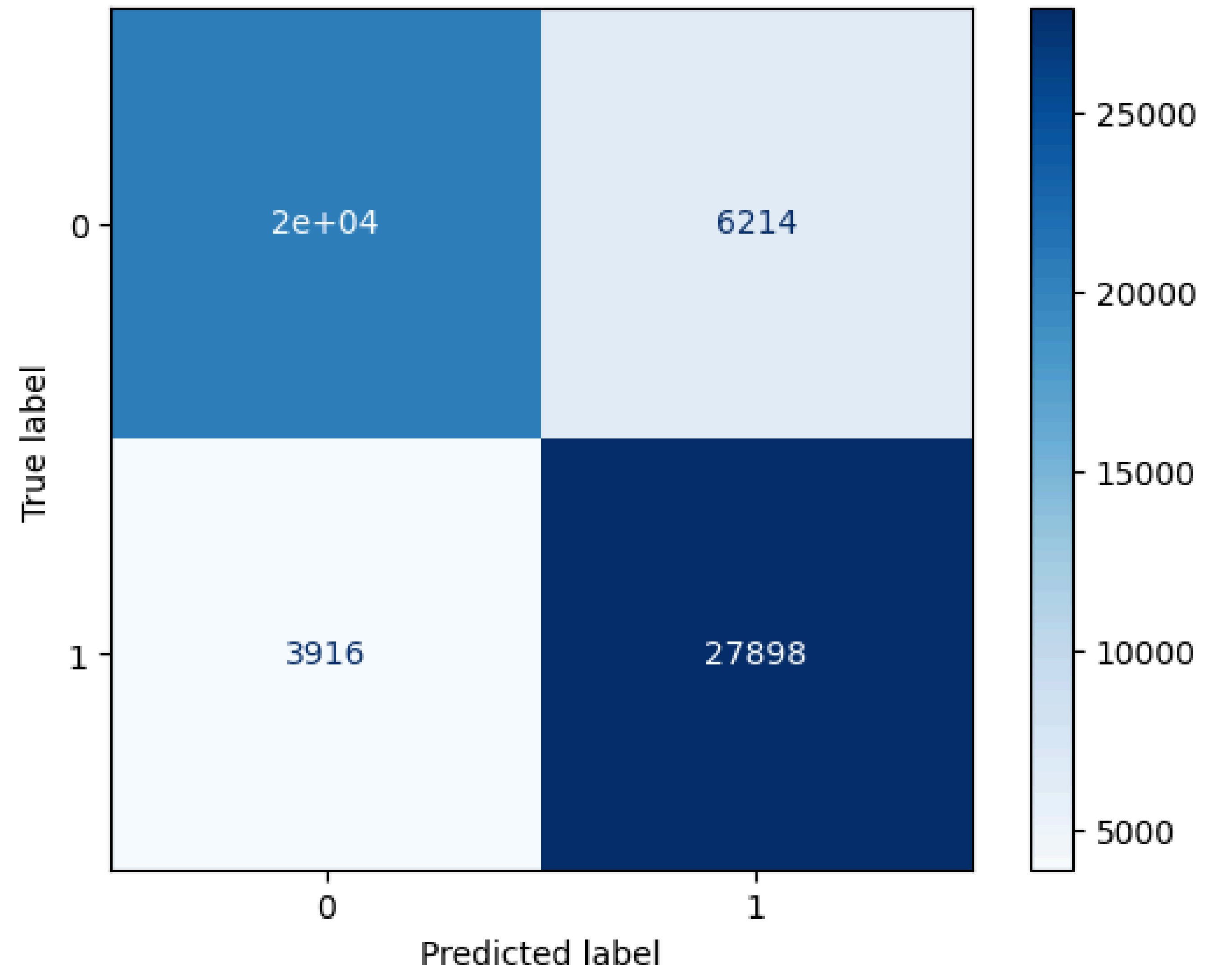
- n_estimators: 100 to 350
- max_depth: None, 10, 20, 30
- min_samples_split: 2, 5, 10
- min_samples_leaf: 1, 2, 4
- max_features: 'sqrt', 'log2'
- bootstrap: True, False

```python
param_dist = {
    'n_estimators': np.arange(100, 400, 50),
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['sqrt', 'log2'],
    'bootstrap': [True, False]
}
```

Best Performance at:

```
 Best Result:
params      {'n_estimators': 300, 'min_samples_split': 2, 'min_samples_leaf': 2, 'max_features': 'sqrt', 'max_depth': 30, 'bootstrap': False}
mean_auc                                                                                                                    0.893464
Name: 11, dtype: object

📁 Results saved to: manual_hyperparameter_results.csv
```

```
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.77      0.80     26704
           1       0.82      0.88      0.85     31814

    accuracy                           0.83     58518
   macro avg       0.83      0.82      0.82     58518
weighted avg       0.83      0.83      0.83     58518

AUC-ROC Score: 0.9025468641538107
```



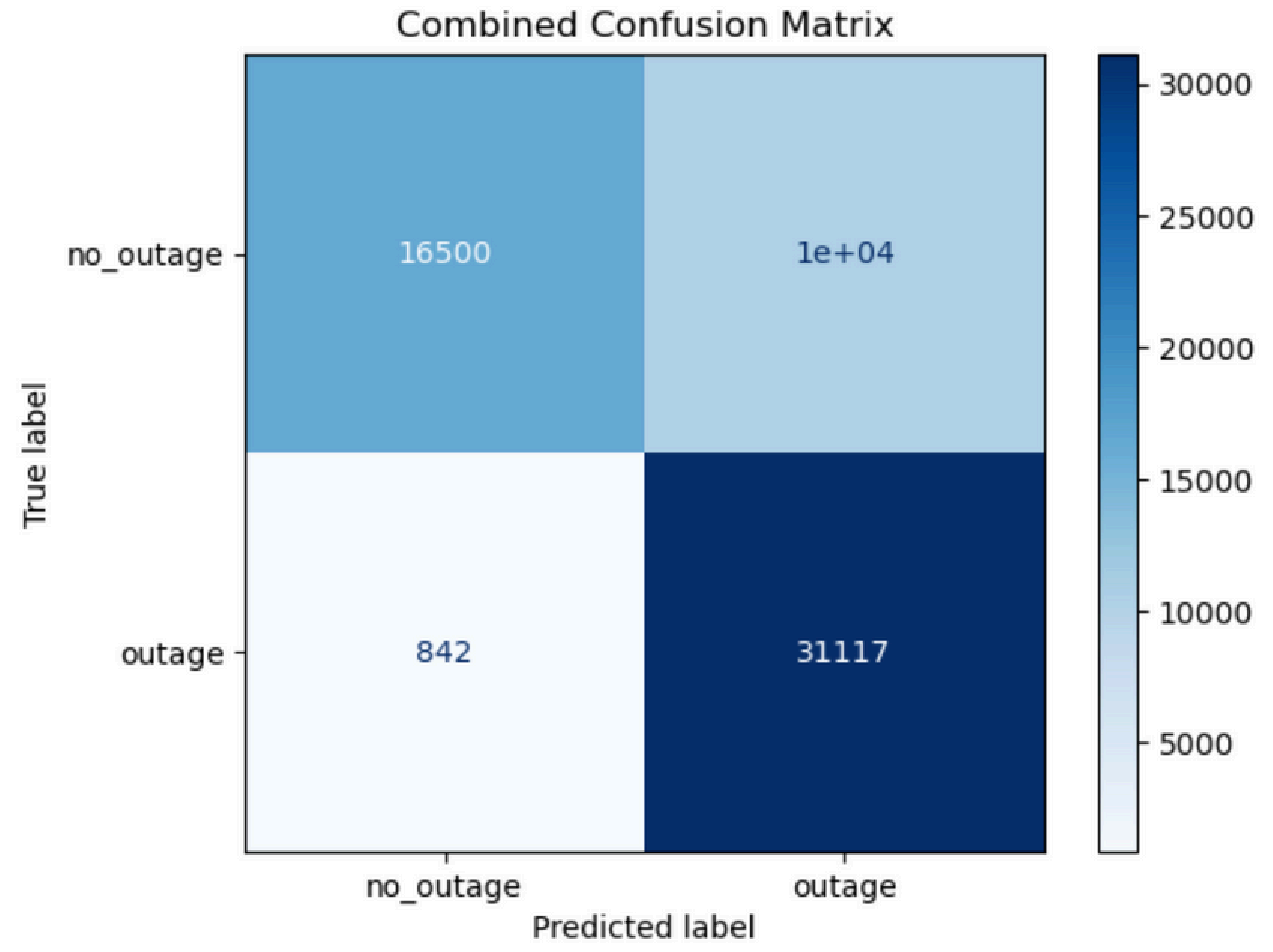
Using manual hyperparameter tuning with randomized sampling and cross-validation, we tested 30 Random Forest configurations. This approach explored a wide parameter space and helped identify the best-performing model with the highest AUC score. The tuned model outperformed the default, offering more accurate and reliable predictions.

Attempt 2: Different Models for different Weather Events

- We our original data by event_type (high_wind, heavy_snow, tornado, thunderstorm, hail) and for each event:
  - we applied SMOTE to training data to account for imbalance.
  - Tuned a separate Random Forest model using RandomizedSearchCV.
- Evaluated individual models and combined the metrics across all events.

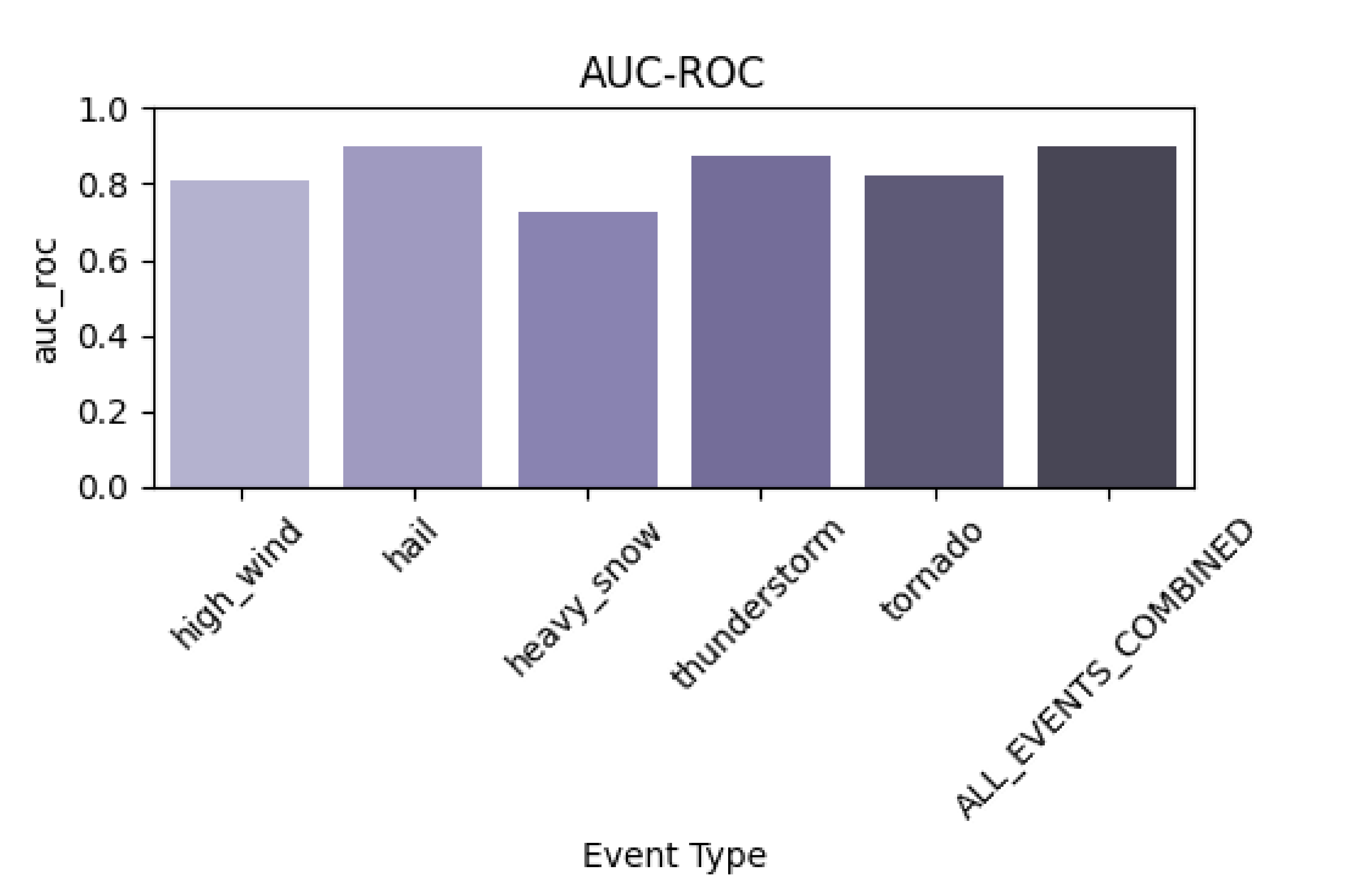
- Results:
  - Per Event performance:
    - High Wind: AUC-ROC: 0.83
    - Heavy Snow: AUC-ROC: 0.819
    - Tornado: AUC-ROC: 0.891
    - Thunderstorm: AUC-ROC: 0.911
    - Hail: AUC-ROC: 0.924
  - Combined: AUC-ROC: 0.92

```
=== COMBINED EVALUATION ===
Combined AUC-ROC Score: 0.92

Combined Classification Report:

              precision    recall  f1-score   support

   no_outage       0.95      0.62      0.75     26817
      outage       0.75      0.97      0.85     31959

    accuracy                           0.81     58776
   macro avg       0.85      0.79      0.80     58776
weighted avg       0.84      0.81      0.80     58776
```



Individual Event Performance

- Separate models were developed for different weather-related events.
- Events caused by high winds and heavy snow showed lower AUC-ROC scores.
- These lower scores indicate reduced performance for these specific hazards.
- As a result, overall model performance declined when such events were included.
- The reason for this was because of lesser data points for these 2 events



Limitations & Deployment Challenges

- Data Imbalance:
  - Very limited samples for heavy snow → poor F1 score even after oversampling
- Geographic Limitation:
  - Model trained only on U.S. data → not generalizable to other regions like India
- Deployment at Plaksha?
  - ❌ Not feasible currently – weather & outage data not available for India
  - ✅ Could work in U.S. using real-time weather forecasts to predict outages
- Scalability Challenge:
  - Scaling would require large, labeled, location-specific datasets and system integration with weather APIs & grid infrastructure

Thank You

State of Art

He et al., 2023 (XGBoost Model for Duration Prediction)
Dataset Size: 50,000 records
Number of Features: 15 features (including weather data, grid information, historical outage data, and duration)
Results:
Accuracy: 87%
Precision: 85%
Recall: 82%
F1-Score: 83%
The model successfully identified long-duration outages, improving restoration time predictions.

Liu et al., 2024 (Random Forest & SVM for Tree-Caused Outage Risk)
Dataset Size: 75,000 records
Number of Features: 12 features (tree types, historical outages, wind speed, storm data, etc.)
Results:
Precision: 90%
Recall: 88%
F1-Score: 89%
AUC-ROC: 0.92
The model effectively predicted tree-caused outages during storms, with high precision in risk classification.

Yilmaz et al., 2023 (Bagging & Logistic Regression for Extreme Weather)
Dataset Size: 60,000 records
Number of Features: 10 features (weather patterns, outage severity, historical outage data, geographical information)
Results:
Precision: 84%
Recall: 80%
F1-Score: 82%
AUC-ROC: 0.87
The model showed good performance in predicting outages caused by extreme weather events.

Mdulansk Project (K-NN for Predicting Outage Causes)
Dataset Size: 25,000 records
Number of Features: 8 features (outage cause, weather data, geographic region, historical outage data)
Results:
Accuracy: 79%
Precision: 77%
Recall: 73%
F1-Score: 75%
The K-NN model identified the causes of outages with reasonable accuracy, though it struggled with rare causes.

Rizvi, 2023 (Deep Learning for Power Outage Prediction)
Dataset Size: 100,000 records
Number of Features: 20 features (including weather data, grid conditions, event history, location, and time of occurrence)
Results:
Accuracy: 91%
Precision: 88%
Recall: 86%
F1-Score: 87%
AUC-ROC: 0.93
The deep learning model outperformed other models, particularly in predicting outages under varying weather conditions.