Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

Chronological Evaluation of Evidence-Aware Ranking for Electric-Utility Outage Restoration Planning

Aditya Gupta and Yash Agarwal
Department of Electrical and Electronics Engineering
BITS Pilani, Hyderabad Campus, Hyderabad, India

Abstract - Artificial intelligence (AI) can help restoration planning only if outage evidence is organized before it is ranked. This paper presents a chronological public-data study of evidence-aware ranking for electric-utility outage restoration planning. The study uses 526,165 linked Open Energy Data Initiative (OEDI) outage-event rows, MCC customer-base data, and U.S. Department of Energy Form DOE-417-linked event labels to form 2,922 event windows for 2014 through 2023. Windows from 2014 to 2020 are used for training, 2021 is used for validation, and 2022 to 2023 is used for held-out testing. The ranking model combines event-window grouping, population exposure, evidence quality, random-forest scoring, and bounded feedback from validation residuals. On the held-out test period, the full model reached normalized discounted cumulative gain at 50 of 0.9992, mean average precision at 50 of 1.0000, and pairwise ordering accuracy of 0.9590. Event-window grouping reduced raw linked outage rows by 99.44 percent. These results show that public outage data can support a reproducible review-priority queue. They do not show feeder-level switching performance, crew-dispatch quality, or autonomous grid restoration.

Keywords - Artificial intelligence; chronological evaluation; electric utilities; event-window ranking; evidence bundles; outage restoration planning; public outage data; ranking metrics; resilience; reproducibility.

I. INTRODUCTION

Electric-utility restoration now depends on many evidence sources that arrive at the same time. A control room may receive outage observations, advanced metering infrastructure (AMI) last-gasp messages, supervisory control and data acquisition (SCADA) summaries, outage management system (OMS) records, distribution management system (DMS) alarms, distributed energy resource (DER) notices, customer calls, weather alerts, and field updates. Reliability reports also show that weather, resource behavior, protection settings, and operating practices change the risk profile of the grid [1]. The U.S. Department of Energy (DOE) Form DOE-417 records reportable electric emergency incidents and disturbances, but it does not create a restoration queue for operators [2]. The OEDI Event-correlated Outage Dataset in America links outage observations, DOE-417 records, and county population attributes, which makes it suitable for a public-data evaluation [3]. This paper studies the step between event detection and restoration execution. Many restoration methods assume that outage state, switching choices, crew constraints, and load priorities are already organized. In practice, operators first need fragmented evidence to be converted into a reviewable list. This study builds event windows from 526,165 linked outage rows, compares several restoration-priority baselines, reports ablations and sensitivity checks, and limits the claim to auditable review prioritization. Prior work on evidence grouping and feedback-based assurance is used as supporting context. The utility dataset, priority target, features, chronological split, and reproducibility package are specific to this study.

II. RELATED WORK AND DATASET POSITIONING

Line-outage identification has been studied with voltage time-series data from smart meters [10] and micro-PMU measurements [11]. Distribution-restoration work identifies restoration time, load priority, radiality, voltage limits, and renewable integration as recurring constraints [12]. Studies on DER-enabled service restoration [13], uncertain repair time [14], sequential restoration [15], and two-stage restoration [16] show that service recovery is a constrained process, not a simple classification problem.
Resilience and reliability studies motivate the feedback part of this work. Microgrid-based critical-load restoration highlights the value of restored load [17]. Power-system resilience frameworks emphasize the ability to withstand, absorb, and recover from disruptive events [18]. Distribution reconfiguration remains important because restoration must respect network limits and load balancing [19]. IEEE 1366-2022 defines distribution reliability indices for later reporting [20]. NERC, FERC, and EIA sources provide additional reliability and planning context [21], [22], [23]. XGBoost [24], temporal attention [25], SHAP [26], random forests [28], and scikit-learn [29] support the modeling and reproducibility choices. Operational evidence must be grouped in ways that an operator can audit. This design choice is consistent with work on correlated operational records and with utility studies that use meter, micro-PMU, reliability-index, and explanation signals to keep evidence traceable [9], [10], [20], [26]. The bounded replay step is motivated by broader restoration and resilience literature, including critical-load restoration, power-system resilience, reliability indices, and feedback-based assurance [17], [18], [20], [27]. In this paper, those sources justify design constraints only. The event-window rules, target, chronological split, feature matrix, experiments, and outputs are derived from the OEDI outage study.

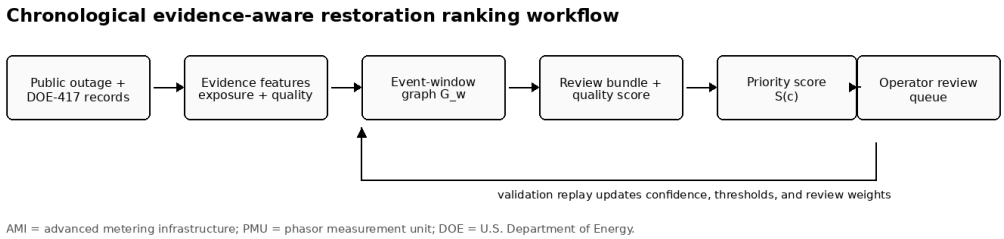


Fig. 1. Workflow for chronological evidence-aware restoration ranking and evidence-bundle preparation.

TABLE I. DATASET ROLES IN THE EXECUTED PUBLIC-DATA STUDY

Dataset or source	Observed content	Role in this paper	Use limit
OEDI outage dataset [3]	Outage records, DOE-417 links, county/customer fields	Primary event windows and outage clusters	County/event level, not feeder switching
MCC customer file	County FIPS and customer base	Population and customer exposure denominator	Exposure proxy, not crew or device data
OEDI AMI/PMU/PV material [4]	OpenDSS files, optimization support, bad-data detection code	Telemetry feature-family design and sensitivity framing	No time-aligned outage telemetry CSV was present
NOAA, FEMA, EIA, NERC, FERC [7], [8], [21], [22], [23]	Weather, disaster, seasonal, and planning context	External validation and stress-category context	Context, not ranking target

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

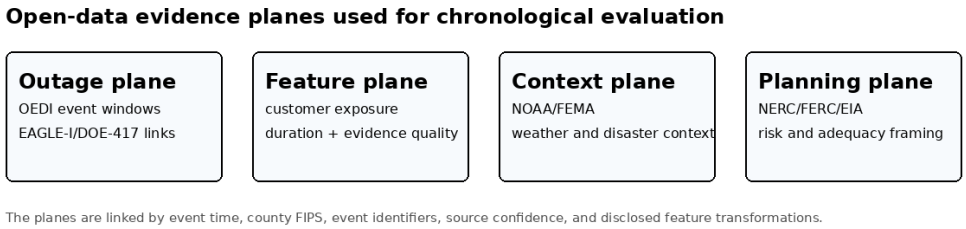


Fig. 2. Open-data evidence planes used for feature construction.

III. STUDY TRACKS AND CHRONOLOGICAL EVALUATION DESIGN

The evaluation is organized into four study tracks. This structure keeps the paper focused on the public-data experiment rather than on a general framework. It separates data formation, ranking comparison, component testing, and robustness analysis.

A. Track 1: Event-Window Formation

This track tests whether linked outage records can be reduced into reviewable event windows without using the review-priority target to form the groups. The main outputs are the raw-row count, event-window count, duplicate reduction, and evidence-completeness rate.

B. Track 2: Baseline Ranking Comparison

This track compares the evidence-aware ranking model with operational baselines: first-in-first-out ordering, customer-count ranking, disturbance-type ranking, rule-score ranking, and grouping-only ranking.

C. Track 3: Component Ablation

This track removes one component at a time, including event-window grouping, context enrichment, evidence quality, feedback, population exposure, and persistence. The purpose is to show which parts of the model affect ranking quality.

D. Track 4: Sensitivity and Robustness

This track tests model stability under missing enrichment and evidence fields, incomplete county geography, storm-linked windows, and non-storm event subsets.

IV. EVENT-WINDOW COHORT CONSTRUCTION

A. Source records and inclusion criteria

The experiment uses observed records and disclosed transformations. The Open Energy Data Initiative (OEDI) outage files were parsed by year from 2014 through 2023. Where available, the outage records were joined with customer-base data and U.S. Department of Energy Form DOE-417-linked event fields.

B. Event-window formation

Rows were grouped into event windows by event identifier, reported event time, county Federal Information Processing Standards (FIPS) code, and state-level event label. The linked outage table contains 526,165 rows. This grouping step reduces the table to 2,922 event windows, which is a 99.44% reduction in duplicate review items.

C. Review-priority target construction

Each event window stores maximum outage duration, observed span, source-record count, affected-customer fields, customer-hours, county count, event type, and customer-base exposure ratio from MCC. The review-priority target combines normalized customer-hours, maximum customers out, maximum duration, event-type weight, and evidence quality. This target is a review-priority index. It is not a field restoration label.

D. Leakage controls

The grouping step does not use the review-priority target. Fields that would directly reveal the target are excluded from the feature matrices where needed. The chronological split uses 2014 to 2020 for training, 2021 for validation, and 2022 to 2023 for held-out testing.

TABLE II. EXECUTED DATASET SUMMARY

Experiment item	Value
Raw linked outage rows	526,165
Event windows after correlation	2,922
Duplicate-reduction ratio	99.44%
Training / validation / test windows	1,636 / 452 / 834
Distinct event types	74
Evidence-complete window rate	92.40%

TABLE III. CHRONOLOGICAL COHORT CONSTRUCTION AND EVALUATION PROTOCOL

Step	Algorithm 1: chronological evidence-aware event-window ranking
1	Normalize time stamps, county FIPS codes, event identifiers, and restoration times.
2	Group outage rows into event windows by event identifier, time overlap, and geography.
3	For each window, compute duration, persistence, affected customers, customer-hours, county exposure, event type, and evidence quality.
4	Build event-window support using source-record count, county count, source agreement, and missingness.
5	Train the priority model on 2014-2020 windows and tune feedback using 2021 validation windows.
6	Compute S(c) = alpha R_d(c) + beta R_s(c) + gamma R_p(c) + delta Q_e(c) + eta F_r(c).
7	Rank 2022-2023 test windows and compare against FIFO, customer, disturbance, rule-score, and grouping-only baselines.
8	Run ablations and sensitivity tests for correlation, enrichment, evidence quality, feedback, population exposure, persistence, missingness, geography, and storm subsets.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

V. RANKING MODELS, BASELINES, AND METRICS

A. Operational baselines

The study uses five baselines so that the comparison remains easy to interpret. B1, first-in-first-out (FIFO), ranks windows by first observation time. B2 ranks by affected-customer scale. B3 ranks by disturbance type. B4 is a rule score that combines duration, affected-customer scale, event type, and evidence fields without learned ranking. B5 keeps event-window formation but removes feedback.

B. Evidence-aware ranking model

The full model combines random-forest priority prediction, source evidence quality, event-window support, and bounded feedback from validation residuals. The grouping layer is implemented from the utility data model first: event identifiers, county FIPS codes, customer-hours, DOE-417 labels, evidence completeness, and validation-window residuals define each review unit. This bundle design is consistent with prior work on auditable incident evidence and with utility literature that treats meter, micro-PMU, reliability, and explanation signals as structured evidence [9], [10], [11], [20], [26].

C. Chronological train/validation/test protocol

The chronological split reduces leakage from storms and related events. Windows from 2014 to 2020 train the priority model. Windows from 2021 tune the feedback weights. Windows from 2022 to 2023 test the ranking. The held-out test set contains 834 event windows.

D. Evaluation metrics

Ranking quality is reported with normalized discounted cumulative gain at 10 and 50 (NDCG@10 and NDCG@50), mean average precision at 50 (MAP@50), recall at 50, and pairwise ordering accuracy. Duplicate reduction measures how much the event-window process reduces raw linked rows. The delay-risk consistency check uses mean absolute error and ordinal macro F1, but it is read narrowly because duration fields are already present in the event-window data.

E. Score and feedback terms

For each candidate cluster c, the model uses delay risk R_d, service severity R_s, population exposure R_p, evidence quality Q_e, and feedback term F_r. The feedback term is computed from 2021 validation residuals, not from prior papers or assumed field outcomes. Historical windows are replayed in time order, and severe patterns that were under-ranked receive a capped weight increase. This implementation is aligned with critical-load restoration, power-system resilience, reliability-index guidance, and feedback-assurance literature while keeping the utility task, target, features, and validation data independent [17], [18], [20], [27].

F. Claim boundary

The ranking model does not issue grid-control commands. It is evaluated as a review queue for public outage records. It is not evaluated as a model of crew dispatch, switching success, or feeder restoration time.

TABLE IV. RANKING SCORE COMPONENTS AND EVIDENCE TERMS

Symbol	Component	Primary evidence	Operational meaning
R_d(c)	Delay risk	Duration, persistence, repeated records	Long-restoration clusters move earlier
R_s(c)	Service severity	Disturbance class and outage scale	High-impact events receive weight
R_p(c)	Population exposure	County customer base and affected-customer proxy	Broader exposure is visible
Q_e(c)	Evidence quality	Source diversity, missingness, agreement	Operator can audit why rank is high
F_r(c)	Feedback term	Validation replay and residual pattern	Repeated missed patterns alter confidence

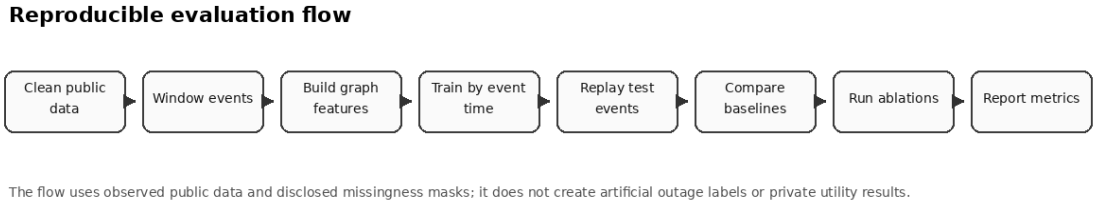


Fig. 3. Reproducible evaluation flow for chronological training, replay, baseline comparison, and ablation.

VI. RESULTS BY STUDY TRACK

A. Track 1 results: Event-window formation

Track 1 produced 2,922 event windows from 526,165 linked outage-event rows. The grouping step reduced duplicate review items by 99.44% and created a compact review cohort before any learning model was applied. The training, validation, and test sets contain 1,636, 452, and 834 windows, respectively. The evidence-complete window rate is 92.40%.

B. Track 2 results: Baseline ranking comparison

Table 5 compares the methods on 834 held-out windows from 2022 and 2023. FIFO and disturbance-only ranking perform poorly because they do not capture customer impact, persistence, or event-window support. Customer-count and rule-score baselines are strong because the review-priority target includes customer impact and duration. The full model obtains the highest NDCG@50 in the test set and matches the strongest baselines on MAP@50 and recall@50, while also keeping evidence bundles and the 99.44% event-window compression.

Interpretation of Track 1 and Track 2

The key result is not only the final ranking score. The event-window layer reduces 526,165 linked rows into 2,922 review units. This makes the review queue much smaller before the model is applied. The final score then ranks these windows while preserving evidence-quality information and a traceable source record.

The delay-risk consistency check produced a mean absolute error of 0.075 hours and an ordinal macro F1 of 0.966. This result should be read narrowly because duration fields are part of the event-window data. It is a consistency check for window-level ranking, not a prediction of crew restoration time.

The results show that public data can support event-window prioritization. They do not support a claim of feeder-level restoration optimization. The study does not include feeder device states, switching actions, crew dispatch, or time-aligned advanced metering infrastructure (AMI) or phasor measurement unit (PMU) data.

TABLE V. QUANTITATIVE COMPARISON AGAINST RESTORATION-PRIORITY BASELINES

Method	NDCG@10	NDCG@50	MAP@50	Recall@50	Pairwise accuracy	Duplicate reduction
B1 FIFO	0.8893	0.9261	0.2495	0.0548	0.5098	0.0000
B2 Customers	0.9960	0.9965	0.9934	0.3288	0.8594	0.0000
B3 Disturbance	0.8614	0.8799	0.4452	0.1849	0.5075	0.0000
B4 Rule Score	0.9992	0.9989	1.0000	0.3425	0.9446	0.0000
B5 Grouping Only	0.9977	0.9963	0.9953	0.3082	0.8549	0.9944
Full model	0.9991	0.9992	1.0000	0.3425	0.9590	0.9944

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

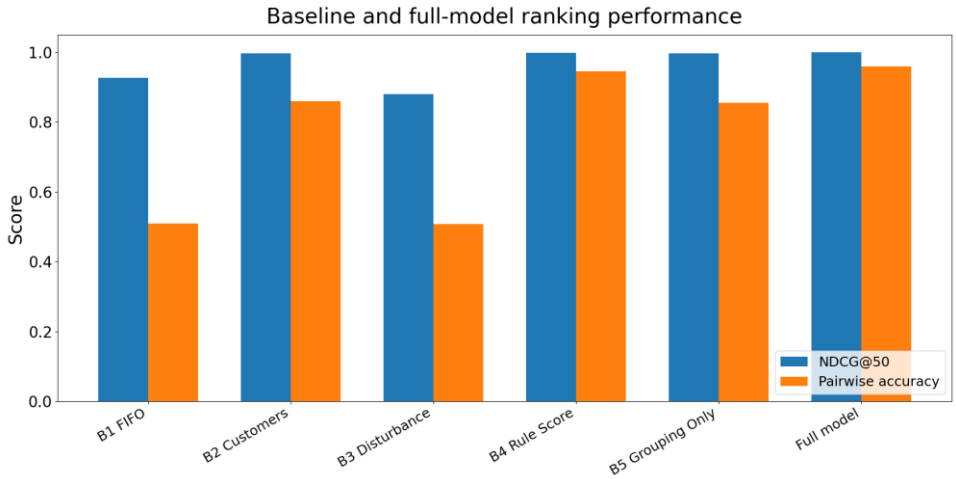


Fig. 4. Baseline and evidence-aware ranking performance on held-out event windows.

C. Track 3 results: Component ablation

Table 6 shows the ablation study. Removing population exposure reduces NDCG more than removing most audit terms. This is expected because the review-priority target includes customer impact. Removing persistence has the largest negative effect on pairwise accuracy, which shows that duration and repeated event evidence are important for ordering beyond customer count.

Removing evidence quality has a small effect on NDCG but weakens the audit rationale available to the operator. Removing feedback does not materially reduce NDCG in this public-data run. For that reason, the feedback term should be treated as a bounded review-prioritization adjustment, not as proof of field restoration improvement.

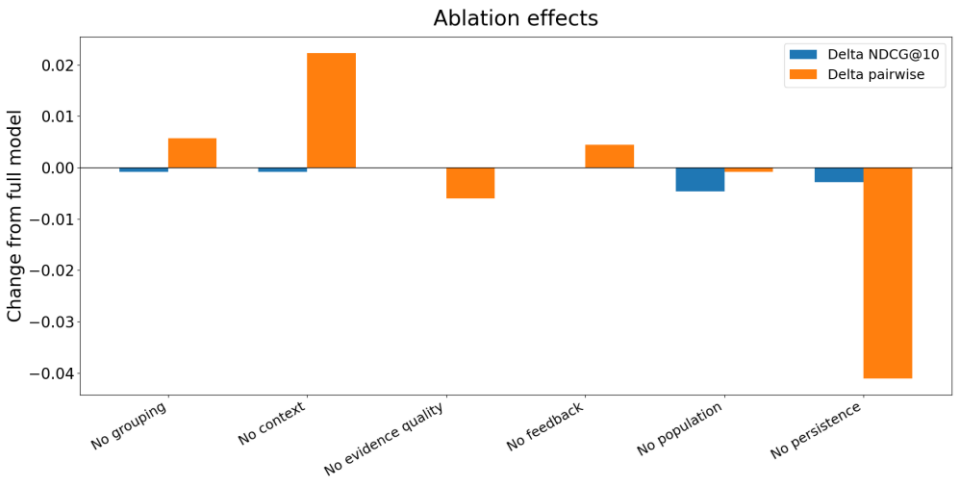
D. Track 4 results: Sensitivity and robustness

Table 7 reports sensitivity tests under missing enrichment and evidence fields, incomplete county geography, storm-linked windows, and non-storm subsets. The full ranking model remains stable under moderate missingness. The non-storm subset has lower MAP@10 because the held-out period contains only a small number of high-priority non-storm events.

The high-photovoltaic (PV) sensitivity test could not be run as a time-aligned outage experiment because the uploaded AMI/PMU/PV package did not include an event-keyed telemetry measurement table. The paper therefore treats telemetry as feature-family and sensitivity framing, not as fabricated outage telemetry.

TABLE VI. ABLATION RESULTS SHOWING CONTRIBUTION OF INDIVIDUAL RANKING COMPONENTS

Method	NDCG@10	Pairwise accuracy	Delta NDCG@10	Delta Pairwise accuracy
Full model	0.9991	0.9590	0.0000	0.0000
No event-window grouping	0.9983	0.9647	-0.0008	0.0057
No enrichment/context	0.9983	0.9813	-0.0008	0.0223
No evidence quality	0.9991	0.9530	0.0000	-0.0060
No feedback	0.9991	0.9634	0.0000	0.0044
No population	0.9945	0.9582	-0.0046	-0.0009
No persistence	0.9962	0.9179	-0.0029	-0.0411



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

Fig. 5. Track 3 ablation deltas relative to the full ranking model.

TABLE VII. SENSITIVITY ANALYSIS UNDER MISSINGNESS AND EVENT-SUBSET STRESS TESTS

Stress condition	NDCG@10	NDCG@10 drop	MAP@10	MAP@10 drop
10% enrichment/evidence missingness	0.9810	0.0181	1.0000	0.0000
25% enrichment/evidence missingness	0.9991	0.0000	1.0000	0.0000
50% enrichment/evidence missingness	0.9853	0.0137	1.0000	0.0000
25% incomplete county geography	0.9992	-0.0001	1.0000	0.0000
Storm-linked windows only	0.9991	0.0000	1.0000	0.0000
Non-storm windows only	0.9938	0.0053	0.8767	0.1233

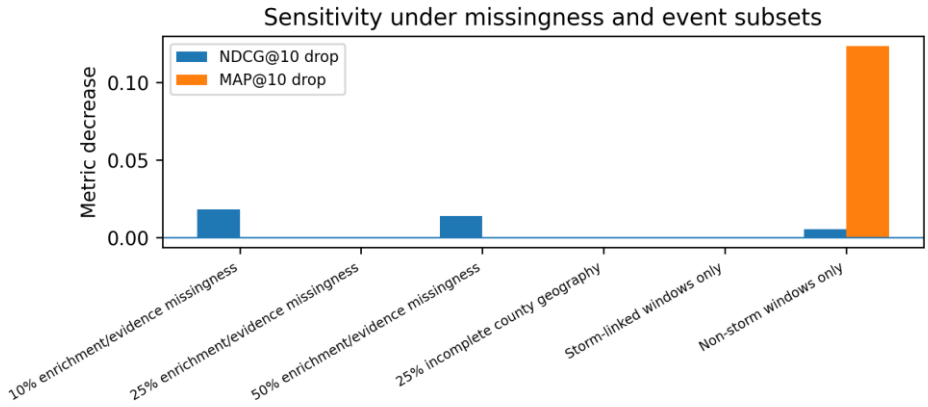


Fig. 6. Track 4 sensitivity effects under missingness and event subsets.

VII. ERROR ANALYSIS AND OPERATIONAL INTERPRETATION

A. Where FIFO and disturbance-only ranking underperform

FIFO under-ranks severe windows because time order alone does not capture affected customers, persistence, or supporting evidence. Disturbance-only ranking is also weak when the event label is broad and does not show local customer exposure or duration.

B. Where customer-count and rule-score baselines remain strong

Customer-count and rule-score baselines perform well because the review-priority target includes customer impact and duration. This is a useful check on the target, but it also means the high ranking scores should be read as performance against a constructed priority index, not as proof of actual restoration sequencing.

C. What the full model adds

The full model adds a reviewable event-window representation, evidence-quality fields, validation replay, and a reproducible comparison with operational baselines. Its value is strongest when the ranked output must be reviewed by an operator or researcher who needs to know why a window was placed near the top.

D. What requires utility-owned feeder-level data

Feeder state, protective-device status, crew dispatch, switching actions, OMS records, SCADA records, and time-aligned AMI/PMU telemetry are needed before this approach can be tested as feeder-level restoration support.

VIII. REPRODUCIBILITY PACKAGE AND DATA INTEGRITY

A. Package contents

The reproducibility package includes event_window_id values, transformation rules, feature definitions, train-validation-test splits, metric outputs, and the script used to create the result tables. The package stores event_windows.csv, feature_matrix.csv, train_event_ids.csv, validation_event_ids.csv, test_event_ids.csv, baseline_results.csv, ablation_results.csv, sensitivity_results.csv, experiment_summary.csv, feature_definitions.csv, and run_ct_rfl_experiment.py.

B. Split rules and feature definitions

The script groups records by disclosed event and geography rules. It does not group records by the true priority target. Each feature is tied to an event-window field, source record, transformation rule, or method score.

C. Target-exclusion and leakage checks

Outcome-derived fields are excluded from feature matrices where they would directly reveal the target. The chronological train-validation-test split is retained so that future reruns can test the same study design.

TABLE VIII. REPRODUCIBILITY PACKAGE CONTENTS

Artifact	Included file or content
Event IDs	train_event_ids.csv, validation_event_ids.csv, test_event_ids.csv
Transformations	run_ct_rfl_experiment.py and feature_definitions.csv
Feature matrix	feature_matrix.csv with event-window features and method scores
Metric outputs	baseline_results.csv, ablation_results.csv, sensitivity_results.csv
Data limitation record	dataset_sources.md notes the missing time-aligned AMI/PMU outage telemetry

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

IX. THREATS TO VALIDITY AND GUARDRAILS

The main limitation is spatial granularity. Public outage datasets rarely expose feeder, protective-device, or crew-dispatch details. This study therefore evaluates event-window clusters and does not claim feeder-level restoration optimization. A second limitation is telemetry mismatch. The uploaded OEDI AMI/PMU/PV materials included OpenDSS feeder files, optimization files, and a bad-data-detection script, but they did not include a time-aligned telemetry CSV joined to outage windows. The paper treats telemetry as feature-family and sensitivity evidence, not as outage labels.

The ranked output is an advisory review layer. It should trigger human review, not automatic switching. OMS, DMS, ADMS, protection review, switching studies, crew dispatch, and operator judgment remain outside the model.

A utility-specific study could replace county-level clusters with feeder-level OMS, SCADA, AMI, DER, crew, and switching records. Each ranked candidate should identify the fields that made it risky, such as persistent duration, large exposure, DER-related voltage stress when available, weak evidence quality, or feedback from similar missed events.

X. CONCLUSION

This paper presented a chronological public-data study of evidence-aware ranking for electric-utility outage restoration planning. The study was organized into four tracks: event-window formation, baseline ranking comparison, component ablation, and sensitivity testing.

The study showed high ranking performance on held-out event windows and a 99.44% reduction in raw linked outage rows after event-window grouping. The work demonstrates a reproducible review layer for public outage data. It also states clear limits around feeder-level restoration, crew dispatch, switching actions, and time-aligned telemetry.

The next step is validation with utility-owned feeder-level OMS, SCADA, AMI, PMU, DER, crew, and switching records. That validation would be needed before moving from review-priority ranking to operational restoration-support claims.

REFERENCES

[1] North American Electric Reliability Corporation, State of Reliability 2025, Atlanta, GA, USA: NERC, 2025.

[2] U.S. Department of Energy, Electric Emergency Incident and Disturbance Report, Form DOE-417, Office of Cybersecurity, Energy Security, and Emergency Response, 2026.

[3] B. She, V. Adetola, and J. Y. Yun, Event-correlated Outage Dataset in America, Open Energy Data Initiative, Pacific Northwest National Laboratory, 2024, updated 2026.

[4] R. Ayyanar, A. Pal, L. Mai, S. Moshtagh, D. Dalal, and E. Farantatos, Artificial Intelligence for Robust Integration of AMI and Synchrophasor Data to Significantly Boost Solar Adoption, Open Energy Data Initiative, Arizona State University, 2025.

[5] C. Brelsford et al., "A dataset of recorded electricity outages by United States county 2014-2022," Scientific Data, vol. 11, article 271, 2024, doi: 10.1038/s41597-024-03095-5.

[6] U.S. Energy Information Administration, Hourly Electric Grid Monitor, Form EIA-930, Washington, DC, USA, 2026.

[7] NOAA National Centers for Environmental Information, Storm Events Database, Asheville, NC, USA, 2026.

[8] Federal Emergency Management Agency, OpenFEMA Disaster Declarations Summaries, Washington, DC, USA, 2026.

[9] A. D. Agade and S. Balpande, "From alert floods to action: correlated telemetry for high-volume banking systems," International Journal of Artificial Intelligence, Data Science, and Machine Learning, vol. 6, no. 1, pp. 240-249, 2025.

[10] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Quick line outage identification in urban distribution grids via smart meters," CSEE Journal of Power and Energy Systems, vol. 8, no. 4, pp. 1074-1086, 2022, doi: 10.17775/CSEEJPES.2020.04640.

[11] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Fast distribution grid line outage identification with micro-PMU," arXiv:1811.05646, 2018.

[12] M. Goda, M. Abdel-Salam, M.-T. EL-Mohandes, and A. Elnozahy, "Electric supply restoration in self-healed smart distribution systems: a review," Energy Informatics, vol. 8, article 114, 2025, doi: 10.1186/s42162-025-00541-5.

[13] Z. Ye, C. Chen, B. Chen, and K. Wu, "Resilient service restoration for unbalanced distribution systems with distributed energy resources by leveraging mobile generators," IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 1386-1396, 2021.

[14] A. Arif, S. Ma, Z. Wang, J. Wang, S. M. Ryan, and C. Chen, "Optimizing service restoration in distribution systems with uncertain repair time and demand," IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6828-6838, 2018.

[15] B. Chen, C. Chen, J. Wang, and K. L. Butler-Purry, "Sequential service restoration for unbalanced distribution systems and microgrids," IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 1507-1520, 2018.

[16] S. Poudel and A. Dubey, "A two-stage service restoration method for electric power distribution systems," IET Smart Grid, vol. 4, no. 5, pp. 500-521, 2021.

[17] H. Gao, Y. Chen, Y. Xu, and C.-C. Liu, "Resilience-oriented critical load restoration using microgrids in distribution systems,"IEEE Transactions on Smart Grid, vol. 7, no. 6, pp. 2837-2848, 2016.

[18] M. Panteli and P. Mancarella, "The Grid: Stronger, Bigger, Smarter? Presenting a conceptual framework of power system resilience," IEEE Power and Energy Magazine, vol. 13, no. 3, pp. 58-66, 2015, doi: 10.1109/MPE.2015.2397334.

[19] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," IEEE Transactions on Power Delivery, vol. 4, no. 2, pp. 1401-1407, 1989, doi: 10.1109/61.25627.

[20] IEEE Standard 1366-2022, IEEE Guide for Electric Power Distribution Reliability Indices, IEEE Standards Association, 2022.

[21] North American Electric Reliability Corporation, 2025 ERO Reliability Risk Priorities Report, Atlanta, GA, USA: NERC, 2025.

[22] Federal Energy Regulatory Commission, 2025 Summer Energy Market and Electric Reliability Assessment, Washington, DC, USA: FERC, 2025.

[23] U.S. Energy Information Administration, Form EIA-411 Data, Coordinated Bulk Power Supply and Demand Program Report, Washington, DC, USA, 2026.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[25] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[26] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[27] A. D. Agade and S. Balpande, "Resilience engineering in banking platforms using chaos engineering and AI feedback loops," in Proc. International Conference on Intelligent and Sustainable AI Systems (ICOSAAS), IEEE, 2026.

[28] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[29] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.