Distribution Event Probability Models Version 4 (DEPM v4) Documentation

Prepared By: PG&E Risk and Data Analytics Team

Revision History

Rev	Date	Comment
0	08/27/2024	Initial Release
1	03/12/2025	Added new cover page

Table of Contents

Executive Summary

The Distribution Event Probability Models (DEPM) are used by PG&E to quantify and categorized failure and ignition events from the overhead distribution network assets. The event probability model suite predicts the relative likelihood of failure and ignition events for distribution assets for a wildfire season.

The DEPM model suite is comprised of a top-level Probability of Ignition model and a supporting suite of Probability of Failure/Outage models. There are two types of Probability of Event Failure/Outage models, Equipment and Contact From Object. Equipment models address failures of physical equipment assets on the distribution grid. Contact From Object models predict failures that result when an external object such as a tree, vehicle, or animal contact a grid asset.

The DEPM supports the Wildfire Distribution Risk Model (WDRM) mission to quantify and categorize wildfire risk from overhead distribution network assets. The relative wildfire risk results produced by the WDRM inform the PG&E Wildfire Mitigation Plan (WMP) as well as specific mitigation program work plans to help realize the company stand, *"Catastrophic wildfires will stop"*.

This document describes DEPM version 4 (v4) development by the PG&E Risk and Data Analytics (RaDA) team and highlights improvement implemented over its predecessor, DEPM v3.

DEPM v4 Improvements

DEPM v4 features several significant improvements to the modeling of asset failure and ignition events:

- Substantial improvements to data quality of asset records, event locations, and event causes resulted in improved model performance.
- Model development training data sets were adjusted for the impact of EPSS.
- Several equipment asset models were upgraded from spatial to asset time-series models, resulting in improved model performance and direct aggregation of risk by asset to circuit segments, which in turn improved relative circuit segment risk values for WDRM v4.
- Several new features were added to the Vegetation models to better account for climate change effects and to improve model performance.
- Introduction of event cause and equipment type interaction terms to the Probability of Ignition given Outage model to improve performance for causal pathways that share underlying characteristics for weather and fuels.

DEPM v4 Event Probability Insights

The improvements implemented for DEPM v4 have resulted in generally improved performance metrics when compared with the v3 model. The newly developed equipment asset models provide more causal insight into outage and ignition events for mitigation planning.

An analysis of the v3 and v4 probability of ignition values revealed that the v4 results provide more relative sensitivity for comparing assets, locations, and circuit segment aggregated values. One contributor to the increased risk sensitivity is that the probability of ignition range for v4 has increased by more than an order of magnitude over the range produced by the v3 model. The increased range resulted from the introduction of the new equipment asset models and improved model data quality. Consequently, the probability of ignition for v4 has a larger influence on circuit segment wildfire risk values and ranks for WDRM v4 than for WDRM v3.

DEPM v4 Event Future Plans

PG&E has historically released new versions of the WDRM, and the supporting DEPM, on an annual cycle. In the future, WDRM releases are to be aligned with the Wildfire Mitigation Plan (WMP) filing schedule rather than annually. WDRM v4 is intended to support PG&E's 2026 WMP.

DEPM v4 delivered several features planned for 2024 and 2025 development. Future event probability models will aim to improve on existing features and capabilities while addressing additional unquantified risk. Some of the expected future development includes:

- Extending the event probability model time horizon beyond the wildfire season to cover the entire calendar year in support of public safety and reliability risk models.
- Development of new Underground Equipment Asset models.
- Additional models will be transitioned from using a spatial modeling to a temporal modeling algorithm to improve performance by accounting for changes in model features over time.
- Continuous improvement efforts to the capture, detail, maintenance, and curation of data sets used for the event probability models.
- Incorporate improved FPI model predictions from Meteorology into the Probability of Ignition model.

1 Introduction

1.1 Event Probability Model Overview

A key component in evaluating the operational risk resulting from PG&E's electrical distribution network is the capability to predict the types and frequency of failure, outage, and ignition events. The Distribution Network Event Probability Models integrate collective knowledge about historical distribution grid events to estimate the likelihood of future events.

The Event Probability Model predictions can be used to support multiple types of risk assessment for planning operations and maintenance activities. The most significant use of the models, currently, is contributing to the determination of Wildfire risk to inform long-term mitigation work planning. The model predictions can also be used to assess other types of risk such as Public Safety and Reliability.

1.2 Event Probability Model Types

The types of probability models produced for predicting future events are:

- Probability of Outage models
  - Equipment Asset models
  - Contact From Object models
- Probability of Ignition (Given Outage) model

1.2.1 Equipment Asset Model Types

Equipment Asset models consider event history and contributing factors to predict failure of specific types of electrical equipment. Each asset model uses a unique set of inputs (covariates) from a pool of asset attributes and environmental conditions. For some assets, unique models (sub-sets), are produced for specific types of failures.

The following Equipment Asset models were produced for v4:

- Capacitor Bank
- Dynamic Protection Device (DPD)
- Fuse
- Primary Conductor
  - Line Slap
  - Wire Down
  - Other
- Secondary Conductor
- Support Structure
  - Electrical Failure
  - Structural Failure
- Switch
- Transformer
  - Leaking
  - Failure
- Voltage Regulator
- Other Equipment

1.2.2 Contact From Object Model Types

Contact From Object models consider event history and contributing factors to predict failure caused by contact from foreign objects with electrical assets. Each contact model uses a unique set of inputs (covariates) from a pool of object attributes and environmental conditions. All contact models provide unique models (sub-sets) for specific types of contact failures.

The following Contact From Object models were developed for v4:

- Animal
  - Bird
  - Squirrel
  - Other
- Third-Party
  - Balloon
  - Vehicle
  - Other
- Vegetation
  - Branch
  - Trunk
  - Other

1.2.3 Probability of Ignition Model

The probability of ignition is the likelihood that an asset-based ignition will occur during a fire season. Probability of Ignition is predicted by the Probability of Ignition Given Outage model using the probability of outage predictions from all of the Asset Equipment and Contact From Object models along with other attributes such as environmental conditions. Fire season probabilities of ignitions are individually predicted for each specific Asset Equipment and Contact From Object Model.

1.3 Wildfire Distribution Risk Model (WDRM)

The WDRM is the primary risk model that PG&E uses to prioritize mitigation work to reduce wildfires initiated by the distribution grid. Mitigation program work planners use the WDRM results, along with their mitigation program work planning tools to develop work plans that systematically reduce wildfire risk while considering constraints such as budget allocation, human and equipment resource capacity, and regulatory commitments.

The WDRM is released periodically by the RaDA team to quantify the risk related to PG&E's overhead distribution equipment assets, which consist of overhead transformers, poles and other parts of support structures, lines (conductors), and line-related equipment such as interrupters. The WDRM provides predictions of wildfire risk that occurs during an annual wildfire season of June 1st through November 30th.

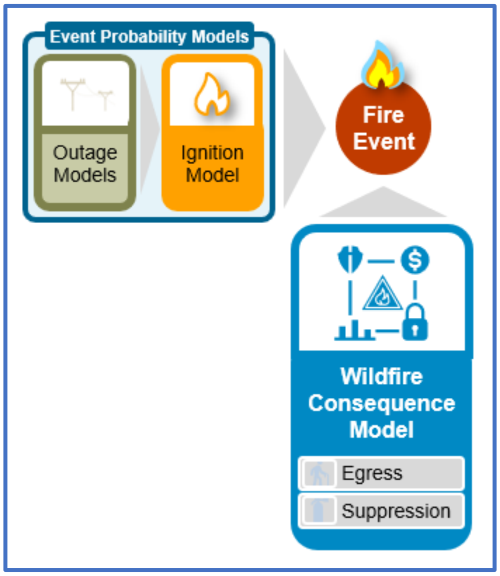

Figure 1 - Wildfire Risk Model

Wildfire risk is calculated from two component values, Probability of Ignition and Wildfire Consequence. Probability of Ignition, p(i), is the annual fire season likelihood that an ignition will originate at the location of an electrical asset. p(i) is predicted by the Event Probability Models described in this document. Wildfire Consequence, WFC, is the expected outcome, or consequence of a fire that ignites at a specific geographic location (of an asset). The Wildfire Consequence Model provides ignition event consequence values and is described in a separate document.

The wildfire risk calculation is:

$$Risk_{loc} = p(i)_{loc} * WFC_{loc} = LoRE_{loc} * CoRE_{loc}$$

where:

CoRE	Consequence of Risk Event, an alternative name for WFC
loc	Asset location expressed as a 100m square pixel within the PG&E service territory
LoRE	Likelihood of Risk Event, an alternative name for p(i)
p(i)	Probability of Ignition
Risk	Wildfire risk
WFC	Wildfire consequence

1.4 Document Content

This document is part of the documentation suite for the Wildfire Distribution Risk Model (WDRM) v4:

- WDRM v4 Documentation
- Wildfire Fire Consequence Model v4 Documentation
- Distribution Network Event Probability Models v4 Documentation (this document)
- RaDA Algorithms and Methodologies

The intent of this document is to present the development, performance, and use of the Distribution Event Probability Models that support the wildfire risk predictions provided by WDRM v4. Source data, modeling approach, and resulting probability predictions are described in detail.

This document seeks to provide the lay reader with a broad understanding of the event probability models. It is not intended to provide comprehensive detailed mathematical or scientific descriptions of all methods used to produce the Event Probability Models.

2 Event Probability Model Improvements

2.1 Version Evolution

The event probability models have been expanded in scope with each release. Generally, with each release new causal models have been introduced that improve the ability of work planners to select the correct mitigation strategy to minimize the outages and failures that can result in ignitions along the distribution grid. Figure 2 provides an overview of how the event probability models have evolved with each version release.

Event Probability Model Evolution

Feature	v1 (2019)	v2 (2021)	v3 (2022)	v4 (2023)
Service Scope	HFTD Tier 2/3	HFTD Tier 2/3	Service Territory	Service Territory
GIS Vintage	2018	2018v/2020c	January 2022	January 2023
Ignitions Event Domain	2015 – 2018	2015 – 2019	2015 – 2020	2015 – 2022
Failures Event Domain	n/a	n/a	2015 – 2021	2015 – 2022
Circuit Segment Aggregation	x	x	Mean Pixel	Risk per Line Mile
Model Compositing	x	x	✓	✓
Asset Models				
Primary Conductor	Pixel-based	Pixel-based	Pixel-based	Asset-based
Secondary Conductor	x	x	Pixel-based	Pixel-based
Support Structure	x	x	Pixel-based	Asset-based
Transformer	x	x	Pixel-based	Asset-based
Voltage Control	x	x	Pixel-based	Asset-based
Capacitor Bank	x	x	x	Asset-based
Switch	x	x	x	Asset-based
DPD	x	x	x	Asset-based
Fuse	x	x	x	Asset-based
Vegetation Models	Pixel-based	Pixel-based	Pixel-based	Pixel-based
LiDAR Data	x	x	✓	✓
with Tree Canopy Density	x	x	x	✓
Tree Health & Wind Direction	x	x	x	✓
Wind Direction	x	x	x	✓
Animal Models	x	x	Pixel-based	Pixel-based
Third Party Models	x	x	Pixel-based	Pixel-based

Figure 2 – Event Probability Model Evolution

2.2 Probability of Failure/Outage Improvements

Several general improvements focused on data pipelines and quality were made to the v4 event probability failure and outage models, including:

- Addition of 2022 events for Failure/Outage and Ignitions.
- Association of historical equipment failure events with the equipment ID of the asset that failed, even if that asset was removed from service. This included a review of the wildfire season events for each time series asset model.
- Enhanced data set reliability through implementation of data pipeline best practices including unit test coverage and setting schema and value data quality expectations.
- Review of the keyword text extraction to identify failure sub drivers based on repair records and field notes. This is particularly important for outage events, where the primary cause and sub cause may not be updated after an initial troubleshooter review.
- Enhanced latitude and longitude locations from vegetation outage inspections for improved fault location of vegetation failure events.

Most asset models have been upgraded to produce asset-based time-series model results rather than spatial grid pixel predictions, enabling the direct aggregation of asset event probabilities and risks to its specific circuit segment for v4. In contrast, v3 asset results were assign to its containing 100m by 100m grid pixel and the asset event probabilities and risk were shared equally with all circuit segments that touched the pixel. This upgrade results in better understanding of relative circuit segment risk for the WDRM. The following models produce asset-specific time-series results for v4:

- Capacitor Bank
- DPD
- Fuse
- Primary Conductor
- Support Structure
- Switch
- Transformer
- Voltage Regulator

Several new features were added to the Vegetation models, accounting for climate change and improving model performance:

- Standardized Precipitation and Evapo-Transpiration Index (SPEI)
- Climatic Water Deficit (CWD)
- California High Hazard Zones from CalFIRE
- Tree Canopy Density from Planet Labs
- Wind Strike Potential
- Improved tree failure location data

2.3 Probability of Ignition Improvements

The key improvements for the Probability of Ignition model include:

- Introduction of event cause and equipment type interaction terms to the Probability of Ignition given Outage model to improve performance for causal pathways that share underlying characteristics for weather and fuels.
- Removing the impact of EPSS activation from the ignitions event training data set.
- Incorporating the expanded set of Probability of Failure/Outage causal models.

3 Event Probability Model Descriptions

3.1 Overview

The following sections describe the development and results for the complete set of v4 probability models. The primary purpose of these models, when combined, is to produce a set of electrical assets probability of ignition estimates for a future fire season for the WDRM.

The event probability models are comprised of a top-level Probability of Ignition model and a supporting suite of Probability of Failure/Outage models. There are two types of Probability of Event Failure/Outage models, Equipment and Contact From Object. Equipment models address failures of physical equipment assets on the distribution grid. Contact From Object models predict failures that result when an external object such as a tree, vehicle, or animal contact a grid asset.

As noted above, the primary purpose driving development of the Event Probability Models is to support the WDRM. However, the Probability of Failure Models can support other risk evaluation needs, such as Reliability and Public Safety Risks.

The v4 event probability models provide event likelihood predictions for distribution network assets as recorded in PG&E's Electric Distribution Geographic Information System (EDGIS) as of January 1st, 2023.

3.2 Model Development Process

All event probability models are developed following an iterative process. A high-level view of the development process is shown in *Figure 3*.

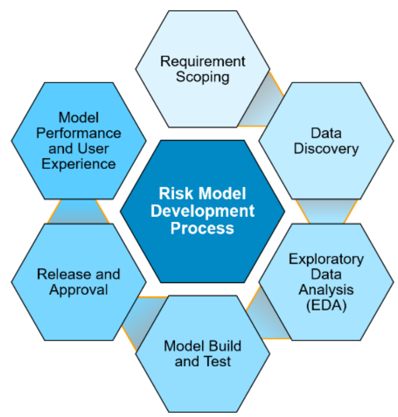

Figure 3 - Model Development Process

Requirement Scoping – Encompasses user feedback, new user requirements, potential performance improvements, and new or desired causal information to set model development goals.

Data Discovery – Researching new data sources, obtaining updates to refresh existing data sources, and cleaning and validating data for potential model use.

Exploratory Data Analysis (EDA) – Data analysis and investigation to determine fitness for modeling.

Model Build and Test – Successive builds of proposed models verified for performance against independent testing data.

Release and Approval – Candidate release models are published to the user community for critique. Viable release candidates are presented through a governance process to gain approval for general release in support of wildfire mitigation work planning.

Model Performance and User Experience – The approved model is monitored for performance against actual events and user feedback is collected to inform the next iteration of model development.

This is a very simplified description of the risk modeling process. Model development is an iterative learning process. As knowledge is gained, model development may require a return to a prior process step.

3.2.1 Model Performance and Feature Metrics

Several metrics are used during the model development process to gauge model quality. Model performance metrics are used to gain insight into how a model performs on data its algorithm hasn't been exposed to during model development. Model feature metrics provide insight into how helpful and influential model feature inputs are to the model outcomes.

Model performance metrics are typically generated by withholding an out-of-sample test data set from the data gathered for model training. After fitting a candidate model using a machine learning algorithm, the candidate is exercised to generate model predictions for the test data set. The candidate model predictions are compared against the actual test data set outcomes to produce performance metrics. The performance metrics commonly used by the RaDA team metrics used for evaluating machine learning models are:

- ROC AUC Curves
- Precision-Recall Curves and Average Precision
- Top 20% Concentration Factor

Features metrics are used to help explain how a machine learning model is influenced by its covariates, why the model predicts a particular set of results, and guides subsequent feature selection for candidate models. Feature metrics are often reviewed with Subject Matter Experts (SMEs) to understand whether or not expected covariates are important to the model and, if not, why not. Conversely, feature metric reviews also carefully consider the importance of unexpected covariates. The most common feature metrics used by the RaDA team include:

- Feature Importance Plots
- SHAP Plots

The model performance and feature metrics are explained in more detail in the RaDA Algorithms and Methodologies document.

3.3 EPSS Impact on Event Modeling

The introduction of Enhanced Powerline Safety Settings (EPSS) as an operational tool for preventing wildfires impacts the target, or training, data sets used to build and calibrate the event probability models. EPSS activation, in general, increases the number of outages for protected conductors while decreasing the number of ignitions. However, as EPSS is not always enabled, the event probability models need to predict event likelihood in the absence of any optional operational control measures. Therefore, the impact of EPSS is removed from target failure and ignition seasonal counts prior to event model calibration.

The EPSS causal impact is inferred from the 2021 EPSS trial period, which was conducted during the 2021 fire season from July 28th through October 23rd, 2021. The trial was conducted on 160 circuits whose protective devices were initially set to the most sensitive default trip setting and were transitioned to less sensitive engineer-specified EPSS settings during the trial. By the start of the 2022 fire season, EPSS settings were available on virtually all circuits in the High Fire Threat District (HFTD), with the settings activated only during periods of elevated ignition risk as determined using PG&E Meteorology's FPI R-score and wind condition forecasts.

The EPSS implementation team estimated program impacts by comparing event rates for EPSS protected circuits from the four years, 2017 – 2020, prior to EPSS enablement and under conditions that would have resulted in

EPSS activation with event rates from when the settings were enabled during the 2022 fire season. Synthetic Difference in Differences (SDID) estimation was applied to event rates for EPSS protected circuits during time periods when EPSS was not enabled in comparison with event rates for non-EPSS protected circuits across the entire wildfire season. The SDID analysis produced estimates of how much EPSS activation increases the rate of outage events and decreases the rate of ignition events.

2022 wildfire season EPSS daily activation records were used to correct the observed event frequency for distribution grid pixel locations where EPSS would likely be activated for future fire seasons. For example, take a set of grid pixels with EPSS protection where:

Variable	Value	Description
$p(e)_{Observed}$	10	Count of observed events, outage or ignition, in a fire season
f_{EPSS}	0.15	Fraction of fire season days with EPSS activated
α_{EPSS}	0.1	Fractional change in events with EPSS activated from SDIS analysis
$p(e)_{Corrected}$	---	Fire season event count corrected for EPSS impact

The EPSS event count correction factor is determined from the observed count, the fraction of EPSS activation days in the fire season, and the fractional change in events with EPSS activated:

$$p(e)_{Observed} = p(e)_{Corrected} * (1 - f_{EPSS}) + p(e)_{Corrected} * (1 + \propto_{EPSS}) * f_{EPSS}$$

Substituting in the example values:

$$10 = p(e)_{Corrected} * (1 - 0.15) + p(e)_{Corrected} * (1 + 0.1) * 0.15$$

$$10 = p(e)_{Corrected} * 0.85 + p(e)_{Corrected} * 0.165 = p(e)_{Corrected} * 1.015$$

$$p(e)_{Corrected} = 9.852$$

The general equation can be simplified to:

$$p(e)_{Corrected} = \frac{p(e)_{Observed}}{1 + f_{EPSS} * \propto_{EPSS}}$$

Again, substituting in the example values:

$$p(e)_{Corrected} = \frac{10}{1 + 0.15 * 0.1} = \frac{10}{1 + 0.015} = \frac{10}{1.015} = 9.852$$

EPSS was implemented first during the 2021 fire season. Therefore, only event counts for the 2021 and 2022 fires seasons are adjusted for EPSS impacts. Event counts for the 2015 through 2020 are used as observed. Therefore, the overall v4 impact for EPSS impact is minor. EPSS impact will become more important for future models as new fire season data is added to the training data sets.

3.4 Training Target Data Sets

The Failures/Outage and Ignition event data used to train and validate the event probability models is produced from the following PG&E data sets:

Table 1 - Event Target Data Set Sources

Data Set	Description	Source
Outages/Forced Outages	Automated detection and documentation of outages that interrupt power to at least one customer.	Integrated Logging Information System (ILIS)
Hazards/Damages	Issues classified as potential hazards or actual equipment damage identified during post-PSPS event inspections.	Post-PSPS Inspection Data
Ignitions	Ignitions with no corresponding ILIS outage are treated as a failure/outage for model development	Electronic Ignitions Database
Emergency A-tags	A-tags with an Outage Information System (OIS) number not included in any other data set above are treated as a failure/outage event for model development.	SAP Notifications

The *Table 1* source event data sets were selected to provide as many failure or fault events as possible representing the modeled failure drivers with accurate event date and time data. The events need an accurate date and time to associate the event with the weather data for the ignition modeling. For this reason, pre-existing failures identified during inspection were not used.

Events used for modeling were filtered using the following criteria:

- Occurred during wildfire season.
- Occurred during or after 2015 when improved ignition recording was initiated.
- Event has a recorded latitude and longitude, a requirement for location-based pixel models.
- Restricted, where known, to overhead assets.
    - Some non-overhead events likely remain in the target data sets due to ambiguous labeling.

For each qualified event, four attributes are extracted from the source data sets to assign the event to the appropriate Equipment Asset or Contact From Object event probability model:

Table 2 - Event Categorization Attributes

Field	Description
Cause	Primary failure cause
Sub-Cause	Failure cause detail
Equipment Type	Asset involved in event
Voltage Category	Primary or secondary system

Table 3 summarizes how categorical attributes are used to assign events to event probability models:

Table 3 - Probability Model Event Assignment

Model	Cause	Sub-Cause	Equipment Type	Voltage Category
Animal-Bird	Animal	Bird	--	--
Animal-Squirrel	Animal	Squirrel	--	--
Animal-Other	Animal	Other	--	--
Capacitor Bank	Equipment	--	Capacitor Bank	--
DPD	Equipment	--	Dynamic Protection Device	--
Fuse	Equipment	--	Fuse	--
Primary Conductor-Line Slap	Equipment	Line Slap	Conductor	Primary
Primary Conductor-Wire Down	Equipment	Wire Down	Conductor	Primary
Primary Conductor-Other	Equipment	Other	Conductor	Primary
Secondary Conductor	Unknown, Environmental, or Equipment	--	Conductor	Secondary
Support Structure-Electrical	Equipment	Tracking or Arcing	Support Structure	--
Support Structure-Equipment	Equipment	Structural or Failure	Support Structure	--
Switch	Equipment	--	Switch	--
Third-Party – Balloon	Third-Party	Balloon	--	--
Third-Party – Vehicle	Third-Party	Vehicle	--	--
Third-Party – Other	Third-Party	Other	--	--
Transformer-Equipment	Equipment	Overloaded or Failure	Transformer	--
Transformer-Leaking	Equipment	Corroded or Leaking	Transformer	--
Vegetation-Branch	Vegetation	Branch	--	--
Vegetation-Trunk	Vegetation	Trunk	--	--
Vegetation-Other	Vegetation	Other	--	--
Voltage Regulator	Equipment	--	Voltage Regulator	--
Other Equipment	Unknown, Environmental, or Equipment	--	Other	--

3.5 Probability of Failure/Outage Models

The probability of failure and outage models predict the likelihood of an electrical grid asset failure or outage during a fire season. This section will describe each of the probability of failure models in detail and has been organized by the two broad types of failure model types, Equipment Asset and Contact From Object, which include:

- Equipment Asset models
  - Capacitor Banks
  - Dynamic Protective Devices (DPD)
  - Fuses
  - Primary Conductors
  - Secondary Conductors
  - Transformers
  - Voltage Regulators
  - Other Equipment
- Contact From Object models
  - Animal
  - Third-Party
  - Vegetation

3.5.1 Equipment Asset Models

The DEPM v4 equipment asset models scale the methodology used for the v3 support structure and transformer models to all but two asset models. The methodology tracks assets over the modeling period using historical snapshots of EDGIS asset data sets. Using historical data sets enables identification of a failed asset even if the asset was removed from service due to the failure. In turn, the attributes of a failed asset can be linked to its failure and used to train an asset model.

The v4 asset-specific models are a significant upgrade over the v3 pixel-based models. A common issue with pixel model data is that a pixel often contained multiple assets, making it difficult to understand precisely which asset was most at risk for failure. The v4 asset-specific models, therefore, are more useful for developing targeted asset management and asset work plans.

Table 4 - Equipment Asset Model Summary

Equipment Asset Event Model	Wildfire Season (mean values)			v4 Probability of Failure		v3 Probability of Failure
	Failures	Ignitions	Ignitions per Failure	Test AUC	Test Average Precision	Test AUC
Capacitor Bank	45	13	0.29	0.52	0.0049	n/a
DPD – Dynamic Protection Device	44	5.3	0.12	0.51	0.0054	n/a
Fuse	299	20.7	0.07	0.67	0.0053	n/a
Primary Conductor – Line Slap	108	7.8	0.07	0.68	0.0001	0.735
Primary Conductor – Wire down	1005	70	0.07	0.76	0.0022	
Primary Conductor – Other	438	23	0.05	0.76	0.0008	
Secondary Conductor	474	21	0.07	0.80	0.0006	0.706
Support Structure – Electrical Failure	467	103	0.22	0.76	0.0044	0.838
Support Structure – Structural Failure	667	26.6	0.04	0.69	0.0010	0.738
Switch	40	5	0.12	0.69	0.0015	n/a
Transformer - Failure	1,960	20.5	0.01	0.77	0.0121	0.632
Transformer - Leaking	56.4	0.1	0.002	0.78	0.0010	0.749
Voltage Regulator	30.6	2.7	0.09	0.62	0.0108	0.882
Other Equipment	7722	29	0.003	0.76	0.0002	0.755

3.5.1.1 Capacitor Bank Model

Parallel installed capacitor banks, also called shunt capacitors, provide power factor correction on longer distribution circuits. They respond to voltage levels by switching in or out of service to boost or buck (reduce) line voltage. Due to the stresses from switching, capacitor Banks have the potential to fail catastrophically with an explosion that leads to an ignition that is both a wildfire and a public safety concern. As documented in *Table 4*, Capacitor Banks have the highest historical ignition to failure rate of all modeled distribution assets.

3.5.1.1.1 Key Developments

Capacitor bank and voltage regulator failure events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, capacitor bank and voltage regulator failure events are modeled separately. Additionally, the v3 spatial model approach was replaced with a time-series model for v4.

The time-series model upgrade was enabled by establishing capacitor bank asset history back to 2016 through an examination of annual snapshots of EDGIS asset data. By examining the historical asset data, failures could be traced back to specific equipment in service at the time of failure and the actual attributes of the failed equipment could in turn be used to train the model.

During model development there were some features that were explored, but not included as covariates in the final model:

- Open tags
- Seasonal meteorology data

3.5.1.1.2 Performance

Capacitor banks were the most challenging equipment asset type to model for v4. Despite training data quality improvements and multiple rounds of new feature experimentation, model performance, as shown in *Figure 4* and *Figure 5*, was very poor, with an AUC of only 0.52. In subsequent discussions with SMEs, it was noted that capacitor banks are regularly inspected every year and therefore do not have a replacement backlog or a maintenance tag backlog. The working theory is that the proactive maintenance of capacitor banks makes the failures more challenging to predict.

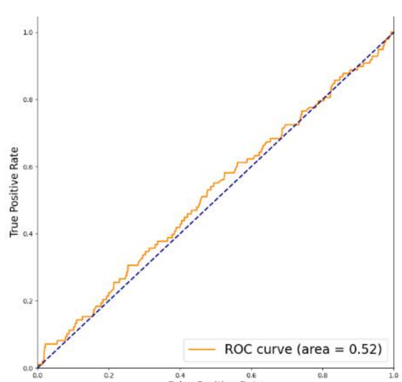

Figure 4 – Capacitor Bank ROC Curve

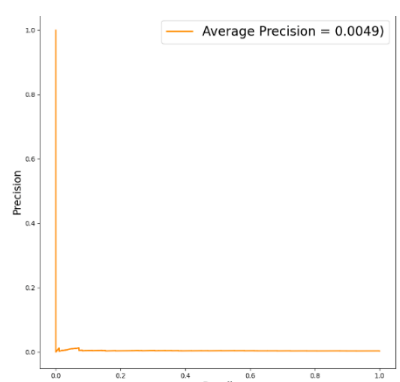

Figure 5 - Capacitor Bank Precision-Recall Curve

3.5.1.1.3 Model Data

The scope of the model was limited to overhead capacitor banks on the distribution system. On average, there were 45 capacitor bank failures and 13 ignitions during wildfire season between 2016 – 2022.

Table 5 – Capacitor Bank Event Summary

Capacitor Bank Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	28	9	16	4
2017	29	10	19	4
2018	28	8	10	3
2019	27	13	10	2
2020	28	6	15	5
2021	38	8	19	4
2022	37	12	9	3
Totals	**215**	**66**	**98**	**25**

3.5.1.1.4 Algorithm Selection

During development, capacitor bank failure models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.1.5 Covariate Selection

Many capacitor bank attributes were tested as features in the model, but none showed major performance gains. *Figure 6* presents the total gains feature importance for the capacitor bank model. Top features include estimated asset age, asset total kVAR capacity, current transformer (CT) rating, supervisory control indicator (remote operation status), switch type, and asset switched off in the winter status.

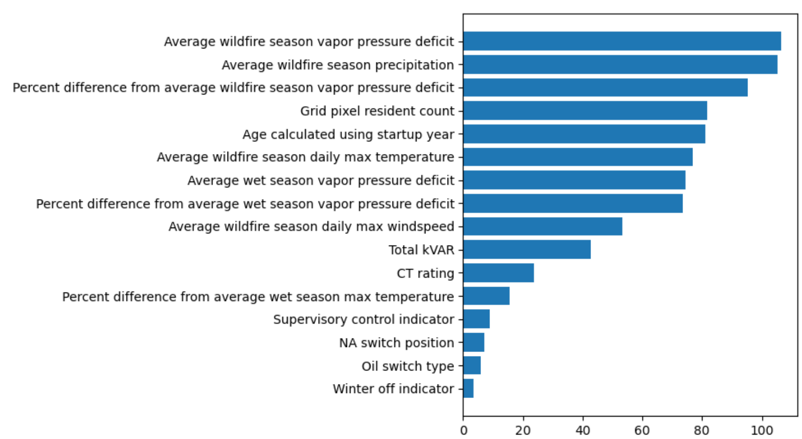

Figure 6 - Capacitor Bank Feature Importance

Figure 7 shows the SHAP plot for the capacitor bank model. Interpreting the plot, model prediction relationships can be understood and validated. For the capacitor banks, higher probability of failure was found to be correlated with:

- High max temperature values
- Low wildfire season precipitation
- High kVAR values
- Low vapor pressure deficits

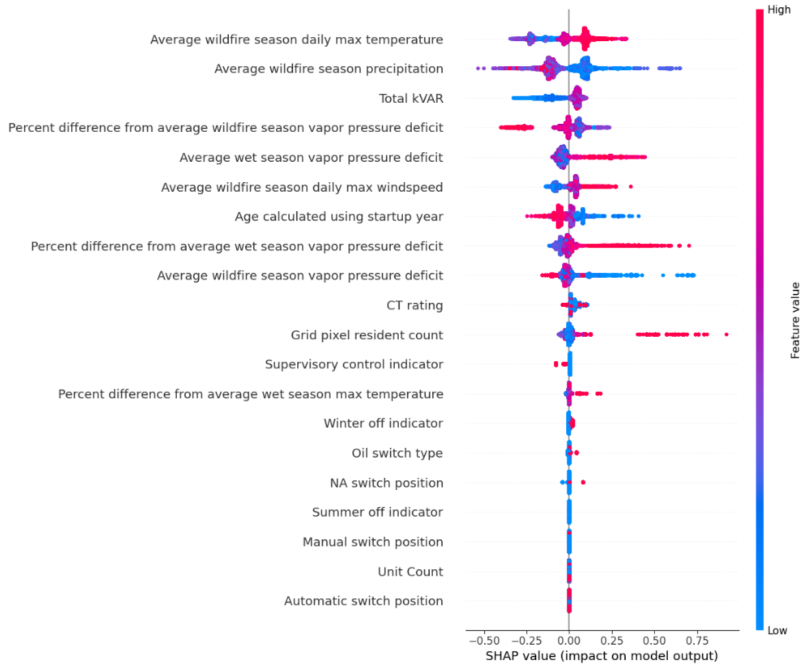

Figure 7 - Capacitor Bank SHAP Plot

3.5.1.1.6 Limitations and Opportunities

Capacitor banks have a very high ignition to failure rate. Therefore, it will be important to improve on the poor performance of the v4 probability of failure prediction model. Since the capacitor bank equipment attributes were not sufficient for predicting capacitor banks failures, new datasets will need to be investigated with SMEs. Switching history, historical loading, and distribution planning load flow (CYME) datasets are recommended for feature engineering in the future. Revisiting failure pathways with SMEs and potential data quality issues for capacitor bank records could be beneficial as well. If the machine learning model continues to perform poorly, a physics-based first principles model will be investigated.

3.5.1.2 DPD Equipment Model

Dynamic Protection Devices (DPD), which include line reclosers and sectionalizers, help protect the distribution system by opening when a fault is detected and attempt to reclose to re-establish service for momentary faults. Examples of DPD equipment failures include when the device malfunctions or the actuator board or controls do not operate as expected in response to a fault.

3.5.1.2.1 Key Developments

DPD equipment, conductor, switch, and fuse events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, DPD equipment was modeled separately from conductors, switches, and fuses. Additionally, the v3 spatial model approach was replaced with a time-series model for v4.

The time-series model upgrade was enabled by establishing DPD asset history back to 2016 through an examination of annual snapshots of EDGIS asset data. By examining the historical asset data, failures could be traced back to specific equipment in service at the time of failure and the actual attributes of the failed equipment could in turn be used to train the model.

During model development there were some features that were explored, but not included as covariates in the final model:

- Open tags
- Seasonal meteorology data

3.5.1.2.2 Performance

DPD Equipment was a challenging equipment asset type to model for v4. Despite training data quality improvements and multiple rounds of new feature experimentation, model performance, as shown in *Figure 8* and *Figure 9*, was very poor, with an AUC of only 0.51.

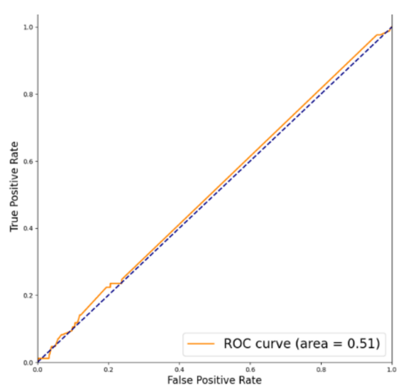

Figure 8 - DPD ROC Curve

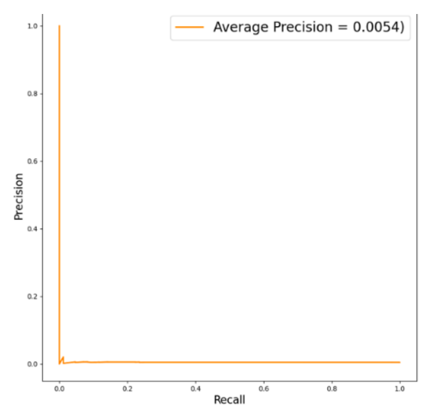

Figure 9 - DPD Precision-Recall Curve

3.5.1.2.3 Model Data

The scope of the model was limited to overhead DPD equipment on the distribution system. On average, there were 44 DPD failures and 5.3 ignitions during wildfire season between 2016 – 2022.

Table 6 - DPD Equipment Event Summary

DPD Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	22	7	17	2
2017	33	2	12	1
2018	26	2	6	1
2019	43	8	18	4
2020	25	1	11	0
2021	34	7	8	2
2022	32	0	13	0
Totals	**215**	**27**	**85**	**10**

3.5.1.2.4 Algorithm Selection

During development DPD failure models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.2.5 Covariate Selection

Many DPD equipment attributes were tested as features in the model, with the relative humidity difference between the average wet season showing the most performance gains. *Figure 10* presents the total gains feature importance for the DPD equipment model. Equipment attributes experimented with during development included estimated asset age, fault duty, continuous current (CC) rating, maximum interrupting current, supervisory control indicator, bypass switch indicator, interrupting medium, device type, and open tags. Of these, the estimated asset age and fault duty were found to be meaningful additions to the model.

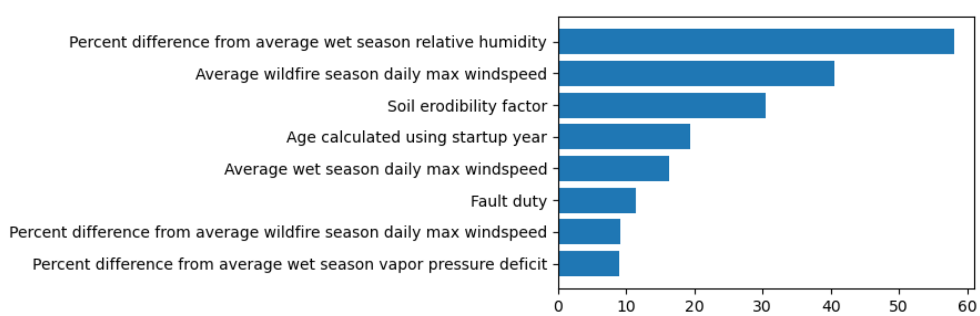

Figure 10 - DPD Feature Importance

Figure 11 shows the SHAP plot for the DPD equipment model. Interpreting the plot, model prediction relationships can be understood and validated. For the DPD equipment, higher probability of failure was found to be correlated with:

- High fire season max windspeed
- High wet season max windspeed
- Soil erodibility
- High estimated asset age

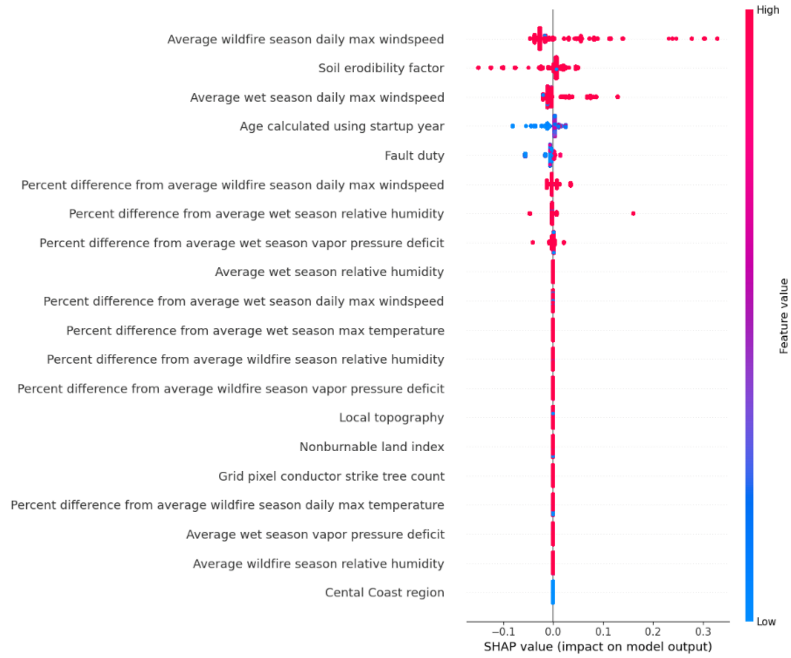

Figure 11 - DPD SHAP Plot

3.5.1.2.6 Limitations and Opportunities

Additional investigation in partnership with SMEs is needed to gain a deeper understanding of what causes DPD equipment to malfunction. This may help to identify new datasets and potential model features. Continuing DPD equipment record data quality improvements, additional failure event validation and consideration of separating out switch and bird guard related events could improve model performance.

3.5.1.3 Fuse Equipment Model

Fuses are electrical safety devices that can help protect the distribution system by preventing the flow of excessive electric current through the system.

3.5.1.3.1 Key Developments

DPD equipment, primary conductor, switch, and fuse events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, fuses were modeled separately from conductors, DPDs, and switches. Additionally, the v3 spatial model approach was replaced with a time-series model for v4.

The time-series model upgrade was enabled by establishing fuse asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, failures could be traced back to specific equipment in service at the time of failure and the actual attributes of the failed equipment could in turn be used to train the model.

During model development there were some features that were explored, but not included as covariates in the final model:

- Open tags
- Seasonal meteorology data

3.5.1.3.2 Performance

The v4 Fuse failure model demonstrated moderately good model performance, as shown in *Figure 12* and *Figure 13*, with an AUC of 0.67.

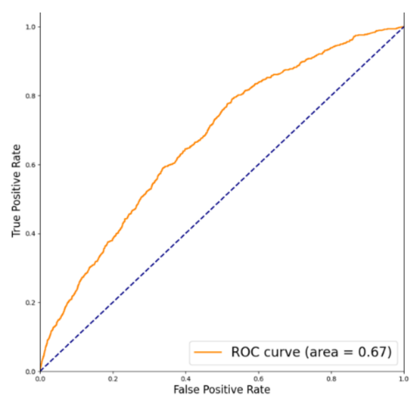

Figure 12 - Fuse ROC Curve

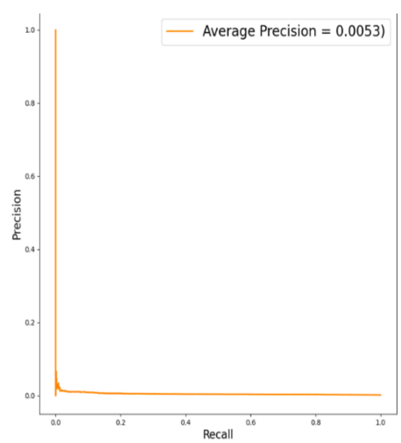

Figure 13 - Fuse Precision-Recall Curve

3.5.1.3.3 Model Data

The scope of the model was limited to overhead fuses on the distribution system. On average, there were 299 fuse failures and 20.7 ignitions during wildfire season between 2016 – 2022.

Table 7 - Fuse Equipment Event Summary

Fuse Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	182	13	90	4
2017	223	26	104	13
2018	191	13	83	5
2019	184	16	93	9
2020	216	9	106	9
2021	203	14	94	5
2022	223	4	101	5
Totals	**1422**	**95**	**671**	**50**

3.5.1.3.4 Algorithm Selection

During development Fuse failure models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.3.5 Covariate Selection

Figure 14 presents the total gains top feature importance for the Fuse model performance. Top features include estimated fuse age, fuse link rating, corrosion zone status, average vapor pressure deficit, average relative humidity, and average operating voltage.

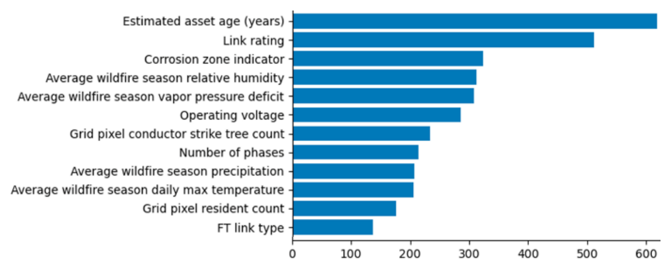

Figure 14 - Fuse Top Feature Importance

Other covariate features that moderately contribute to the Fuse model include:

- Percent difference from average wildfire season relative humidity
- Average wet season relative humidity
- Local topography
- Percent difference from average wet season max temperature
- Soil erodibility factor
- Percent difference from average wildfire season daily max temperature
- Percent difference from average wildfire season daily max windspeed
- Percent difference from average wildfire season daily vapor pressure deficit
- Average wet season vapor pressure deficit
- Percent difference from average wet season relative humidity
- Percent difference from average wet season daily max windspeed
- HFTD
- Complex device indicator
- Bypass switch indicator
- S&C manufacturer
- Percent difference from average wet season vapor pressure deficit

Figure 15 shows the SHAP plot for the Fuse model. Interpreting the plot, model prediction relationships can be understood and validated. For the Fuse equipment, higher probability of failure was found to be correlated with:

- High estimated fuse age
- Low link rating
- In corrosion zone
- Higher number of phases

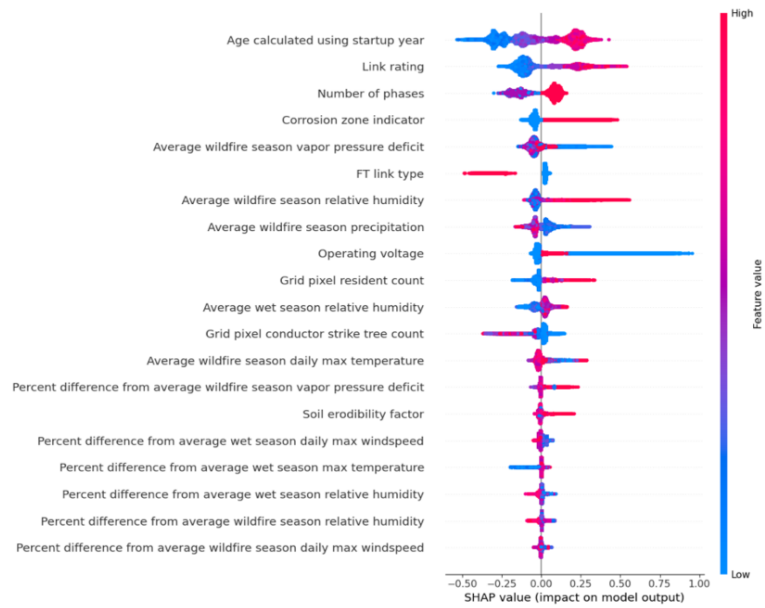

Figure 15 - Fuse SHAP Plot

3.5.1.3.6 Limitations and Opportunities

Future model development is expected to focus on SME collaboration to better understand why link rating and fuse age are so helpful for predicting fuse failures.

During v4 development, it was observed that a notable portion of failed fuses were not mapped in EDGIS until their year of failure. This observation will need to be followed up with EDGIS personnel to understand the current maturity of mapping fuses for the distribution system.

3.5.1.4 Primary Overhead Conductor Models

Primary overhead conductors play a crucial role in delivering electricity safely and reliably to customers. The primary conductors are the system spans of conductor phases that transport electricity from substations to distribution transformers.

Primary overhead conductor, DPD equipment, switch, and fuse events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, Primary overhead conductor failures were modeled separately from fuses, DPDs, and switches. The v3 spatial model approach was replaced with a time-series model for v4 that estimates the likelihood of failure at the nearest pole for the 2023 wildfire season. In addition, the primary overhead conductor failures were separated and developed into three distinct causal models for v4:

- Line Slap – for line slap event failures
- Wire Down – for line to ground failures
- Other – jumper, riser and unknown cause failures

3.5.1.4.1 Primary Overhead Conductor – Line Slap

Primary overhead conductor line slap failures can occur from wind or fault current events that result in two conductor phases becoming physically close enough to cause arcing or sparking which could result in an ignition event. A line slap event can also cause annealing on the conductor which accelerates conductor deterioration and increase the future likelihood of a wire down event.

3.5.1.4.1.1 Key Developments

The Primary Overhead Conductor – Line Slap model was developed as a time-series model for v4, replacing the v3 spatial model for combined primary overhead conduct, DPD, fuse, and switch events.

The time-series model upgrade was enabled by establishing primary overhead conductor asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, line slap failures could be traced back to the nearest pole in service at the time of failure. Tracing conductor failures directly to conductor spans was infeasible as historically failure events are typically recorded at the pole.

Several features were explored during model development, including:

- CYME software fault duty analysis
- Observed splices
- Covered conductor status
- LiDAR span data
- Output from the Finite Element Analysis (FEA) model developed by PG&E's Applied Technology Services (ATS) covering:
  - Wind scenarios
  - Fault current scenarios
- Open repair tags
- Seasonal meteorology data

3.5.1.4.1.2 Performance

The new v4 line slap failure model demonstrated moderately good model performance, as shown in *Figure 16* and *Figure 17*, with an AUC of 0.68.

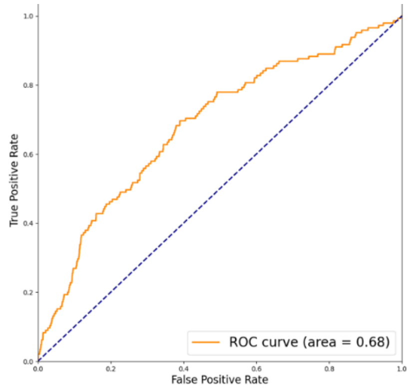

Figure 16 - Line Slap ROC Curve

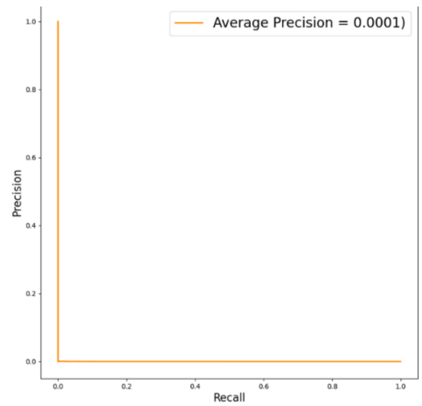

Figure 17 – Line Slap Precision-Recall Curve

3.5.1.4.1.3 Model Data

The scope of the model was limited to overhead conductor line slap events on the distribution system. On average, there were 108 line slap failures and 7.8 ignitions during wildfire season between 2019 – 2022.

Table 8 - Primary Conductor - Line Slap Event Counts

Line Slap Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2019	70	3	47	0
2020	57	5	32	0
2021	107	15	45	5
2022	52	2	21	1
Totals	**286**	**25**	**145**	**6**

3.5.1.4.1.4 Algorithm Selection

During development Primary Overhead Conductor – Line Slap models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.4.1.5 Covariate Selection

Figure 18 presents the total gains top feature importance for the Primary Conductor – Line Slap model performance. Top features include conductor count, conductor length, max conductor weight, minimum conductor cross sectional area, pole height, and wind speed.

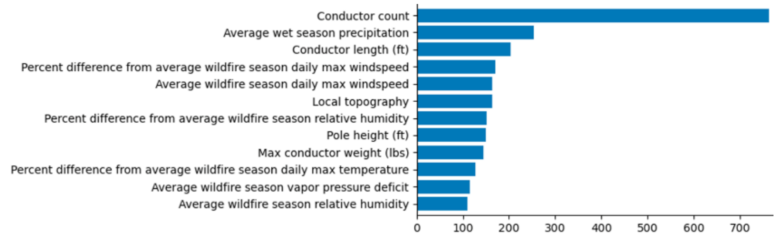

Figure 18 - Primary Condcutor - Line Slap Top Feature Importance

Other covariate features that moderately contribute to the Line Slap model include:

- Grid pixel resident count
- Percent difference from average wet season daily max windspeed
- Percent difference from average wet season vapor pressure deficit
- Average wildfire season daily max temperature
- Percent difference from average wildfire season vapor pressure deficit
- ATS FEA 3000A fault line slap risk
- Average wet season vapor pressure deficit
- Conductor max sag
- Min conductor cross section area
- Grid pixel conductor strike tree count
- Max conductor ultimate strength
- Soil erodibility factor
- Percent difference from average wet season max temperature
- Asset age
- Conductor spacing
- Percent difference from average wet season relative humidity
- Pole elevation
- Average wet season relative humidity
- ATS FEA 60-90 mph wind line slap risk
- Operating voltage
- CYME software fault duty analysis
- Triangular construction type

Figure 19 shows the SHAP plot for the Line Slap model. Interpreting the plot, model prediction relationships can be understood and validated. For the Line Slap cause, higher probability of failure was found to be correlated with:

- Lower conductor phase distances during high winds (FEA prediction)
- Higher conductor count
- Higher conductor max sag
- Higher conductor length
- Pole height
- Higher wet season precipitation

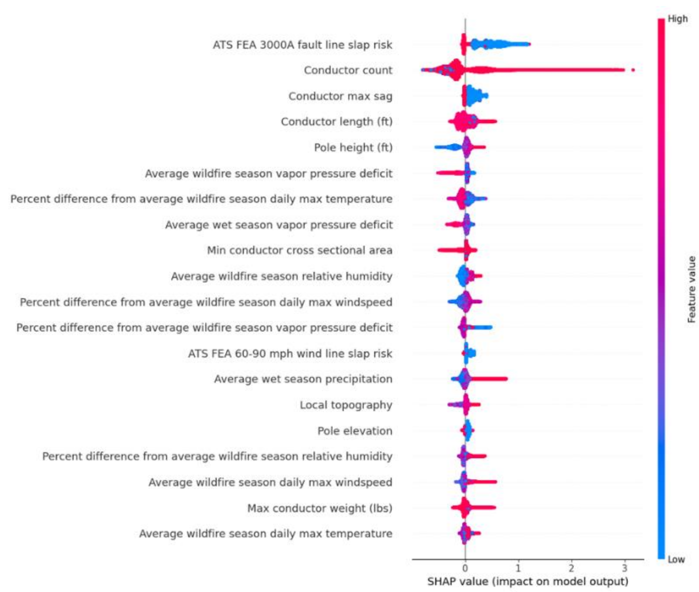

Figure 19 - Primary Conductor - Line Slap SHAP Plot

3.5.1.4.1.6 Limitations and Opportunities

Several initiatives could be pursued to improve the Line Slap model performance for future releases. Several new datasets were acquired during line slap model development. Therefore, another round of data cleanup and feature engineering would likely yield performance gains. Likewise, further review of historical primary conductor event records could identify additional line slap events to use for model training. Identifying additional training events could have a significant performance impact as the line slap event sample size is small in comparison to other primary conductor failure modes. Finally, additional feature engineering with meteorological data, especially wind gusts, could be explored.

3.5.1.4.2 Primary Overhead Conductor – Wire Down

Primary overhead conductor wire down failures are events where a primary wire has fallen or is no longer properly secured. Wire down events can be caused by equipment issues such as the improper installation of a conductor, a connector, or a splice and be degradation issues such as corrosion.

Note that object contacts that result in a wire down event are *not* included in the primary overhead conductor wire down model. All contact events are assigned to one of the Contact From Object models as described in Section *3.5.2*.

3.5.1.4.2.1 Key Developments

The Primary Overhead Conductor – Wire Down model was developed as a time-series model for v4, replacing the v3 spatial model for combined primary overhead conductor, DPD, fuse, and switch events.

The time-series model upgrade was enabled by establishing primary overhead conductor asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, wire down failures could be traced back to the nearest pole in service at the time of failure. Tracing conductor failures directly to conductor spans was infeasible as historically failure events are typically recorded at the pole.

Several features were explored during model development, including:

- CYME software fault duty analysis
- Observed splices
- Covered conductor status
- LiDAR span data
- Output from the Finite Element Analysis (FEA) model developed by PG&E's Applied Technology Services (ATS) covering:
  - Wind scenarios
  - Fault current scenarios
- Open repair tags
- Seasonal meteorology data

3.5.1.4.2.2 Performance

The new v4 wire down failure model demonstrated very good model performance, as shown in *Figure 20* and *Figure 21*, with an AUC of 0.76.

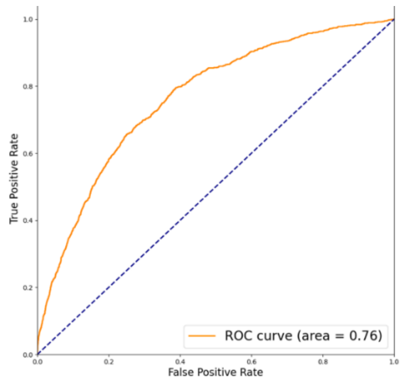

Figure 20 - Wire Down ROC Curve

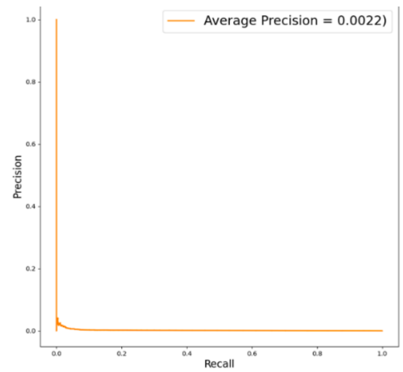

Figure 21 - Wire Down Precision-Recall Curve

3.5.1.4.2.3 Model Data

The scope of the model was limited to overhead conductor wired down events on the distribution system that were not caused by object contact. On average, there were 1005 wire down failures and 70 ignitions during wildfire season between 2019 – 2022.

Table 9 - Primary Conductor - Wire Down Event Counts

Wire Down Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2019	687	42	224	18
2020	649	59	211	18
2021	887	54	311	23
2022	799	50	250	17
Totals	**3022**	**205**	**996**	**76**

3.5.1.4.2.4 Algorithm Selection

Primary conductor wire down events were modeled using SparkML's Random Forest, Gradient Boosted Trees and the spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.4.2.5 Covariate Selection

Figure 22 presents the total gains top feature importance for the Primary Conductor – Wire Down model performance. Top features include conductor count, non-burnable land index, conductor length, max conductor weight, minimum conductor cross sectional area, and height.

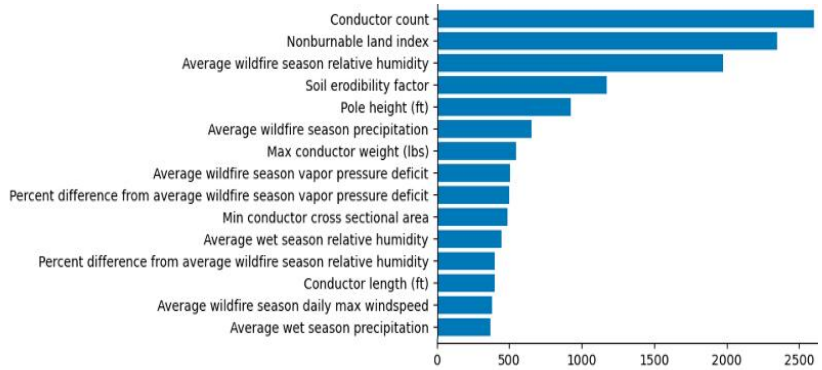

Figure 22 - Primary Conductor - Wire Down Top Feature Importance

Other covariate features that moderately contribute to the Wire Down model include:

- Percent difference from average wildfire season daily max temperature
- Average wet season vapor pressure deficit
- Asset age
- Average wildfire season daily max temperature
- Average wet season daily max temperature
- Coastal indicator
- Percent difference from average wildfire season daily max windspeed
- Max conductor ultimate strength
- Grid pixel conductor strike tree count
- Corrosion zone indicator
- ATS FEA 3000A fault line slap risk
- Percent difference from average wet season max temperature
- Conductor max sag
- Percent difference from average wet season relative humidity
- Percent difference from average wet season daily max windspeed
- Local topography
- Percent difference from average wet season vapor pressure deficit
- CYME software fault duty analysis
- ATS FEA 60-90 mph wind line slap risk
- Operating voltage
- Open tag indicator
- Pole elevation
- Splice observed indicator

Figure 23 shows the SHAP plot for the Wire Down model. Interpreting the plot, model prediction relationships can be understood and validated. For the wire down cause, higher probability of failure was found to be correlated with:

- Higher conductor counts
- Fire season humidity, probably as a proxy for coastal areas linked to corrosion.
- Higher pole heights
- Older conductor ages

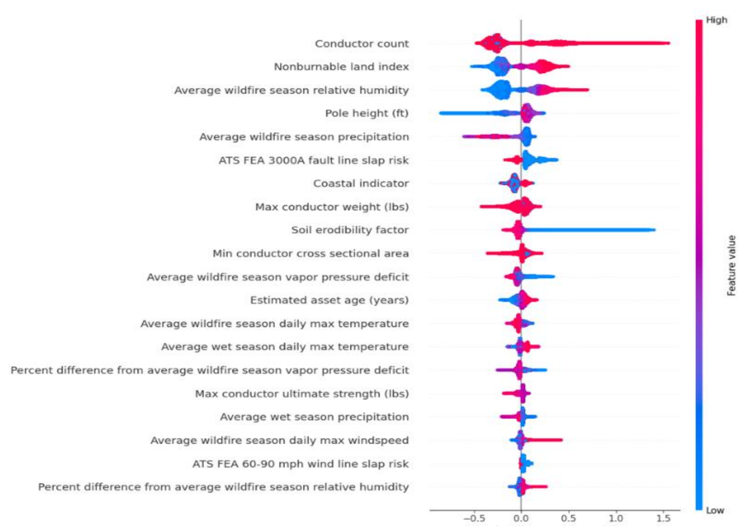

Figure 23 - Primary Conductor - Wire Down SHAP Plot

3.5.1.4.2.6 Limitations and Opportunities

Several initiatives could be pursued to improve the Wire Down model performance for future releases. Several new datasets were acquired during wire down model development. Therefore, another round of data cleanup and feature engineering would likely yield additional performance gains. Likewise, further review of historical primary conductor event records could identify additional wire down events to use for model training. Identifying additional training events could help improve model precision. Finally, additional feature engineering with meteorological data, especially identify areas prone to corrosion, could be explored.

3.5.1.4.3 Primary Conductor – Other

The Primary Conductor – Other model is developed using conductor failures that are not classified as line slap or wired down. Typically, these events include equipment issues related to risers and jumpers as well as conductor-related incidents where the source of failure may have been difficult to identify.

3.5.1.4.3.1 Key Developments

The Primary Overhead Conductor – Other model was developed as a time-series model for v4, replacing the v3 spatial model for combined primary overhead conductor, DPD, fuse, and switch events.

The time-series model upgrade was enabled by establishing primary overhead conductor asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, wire down failures could be traced back to the nearest pole in service at the time of failure. Tracing conductor failures directly to conductor spans was infeasible as historically failure events are typically recorded at the pole.

Several features were explored during model development, including:

- CYME software fault duty analysis
- Observed splices
- Covered conductor status
- LiDAR span data
- Output from the Finite Element Analysis (FEA) model developed by PG&E's Applied Technology Services (ATS) covering:
    - Wind scenarios
    - Fault current scenarios
- Open repair tags
- Seasonal meteorology data

3.5.1.4.3.2 Performance

The new v4 other failure model demonstrated very good model performance, as shown in *Figure 24* and *Figure 25*, with an AUC of 0.76.

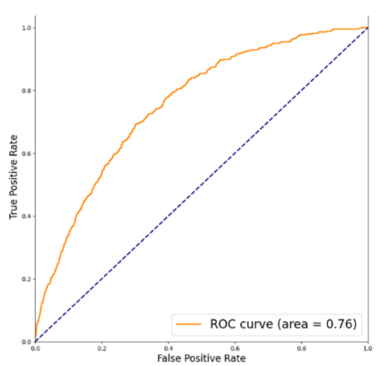

Figure 24 - Primary Conductor - Other ROC Curve

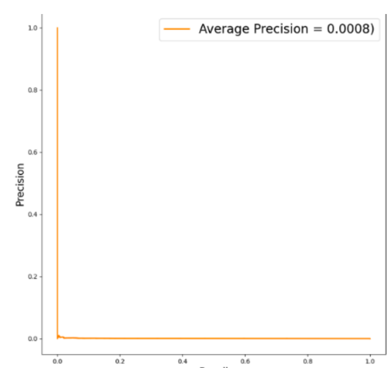

Figure 25 - Primary Conductor - Other Precision-Recall Curve

3.5.1.4.3.3 Model Data

The scope of the model was limited to primary overhead conductor events on the distribution system that were not included in either the Line Slap or Wire Down models. On average, there were 438 other event failures and 23 ignitions during wildfire season between 2019 – 2022.

Table 10 - Primary Conductor - Other Equipment Event Counts

Other Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2019	296	20	93	6
2020	325	25	111	4
2021	369	14	113	5
2022	323	13	121	5
Totals	**1313**	**72**	**438**	**20**

3.5.1.4.3.4 Algorithm Selection

Primary conductor wire down events were modeled using SparkML's Random Forest, Gradient Boosted Trees and the spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.4.3.5 Covariate Selection

Figure 26 presents the total gains top feature importance for the Primary Conductor – Other model performance. Top features include conductor count, non-burnable land index, conductor length, soil erodibility factor, humidity, precipitation, and pole height.

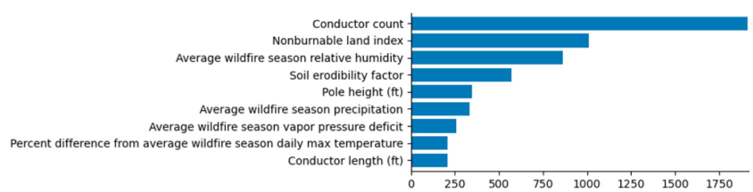

Figure 26 - Primary Conductor - Other Top Feature Importance

Other covariate features that moderately contribute to the Primary Conductor – Other model include:

- Max conductor weight
- Percent difference from average wildfire season vapor pressure deficit
- Average wildfire season daily max windspeed
- Average wet season precipitation
- Asset age
- Percent difference from average wet season relative humidity
- Coastal indicator
- Percent difference from average wet season vapor pressure deficit
- Grid pixel conductor strike tree count
- Average wet season relative humidity
- Average wildfire season daily max temperature
- Percent difference from average wet season max temperature
- Percent difference from average wildfire season daily max windspeed
- Average wet season vapor pressure deficit
- Percent difference from average wildfire season relative humidity
- Local topography
- Percent difference from average wet season daily max windspeed
- Conductor max sag
- Min conductor cross sectional area
- Average wet season daily max temperature
- Triangular construction type
- CYME software fault duty analysis
- Neutral indicator

Figure 27 shows the SHAP plot for the Primary Conductor - Other model. Interpreting the plot, model prediction relationships can be understood and validated. For the other conductor cause, higher probability of failure was found to be correlated with:

- Higher conductor counts
- Non-burnable index, probably as a proxy for urban/non-urban areas
- Fire season humidity, probably as a proxy for coastal areas linked to corrosion.
- Coastal indicator
- Older conductor ages
- Lower fire season vapor pressure deficits

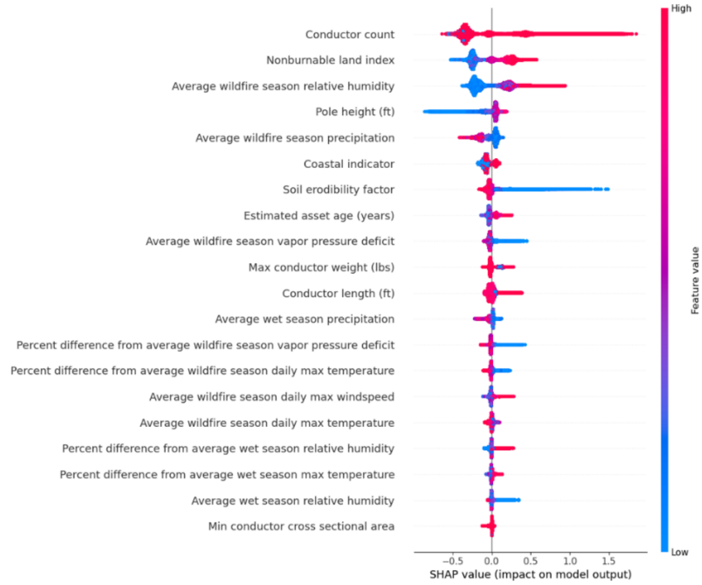

Figure 27 - Primary Conductor - Other SHAP PLot

3.5.1.4.3.6 Limitations and Opportunities

Several initiatives could be pursued to improve the Primary Conductor - Other model performance for future releases. Several new datasets were acquired during model development. Therefore, another round of data cleanup and feature engineering would likely yield performance gains. Likewise, further review of historical primary conductor event records could identify and remove line slap and wire down events to use for model training. Cleaning out line slap and wire down events could in turn help improve Other model performance. Finally, additional feature engineering with meteorological data, especially identify areas prone to corrosion, could be explored.

3.5.1.5 Secondary Conductor Model

The Secondary Conductor equipment model uses historical failure events and spatial environmental data to model future fire season outage probabilities of low voltage secondary conductor assets.

3.5.1.5.1 Key Developments

The Secondary Conductor v4 model is a refresh of the v3 model. The only significant update was the addition of 2023 failures to the event training dataset.

3.5.1.5.2 Performance

The v4 Secondary Conductor outage model demonstrated good model performance with an AUC of 0.78 as shown in *Figure 28*. This was an improvement from the v3 AUC of 0.70

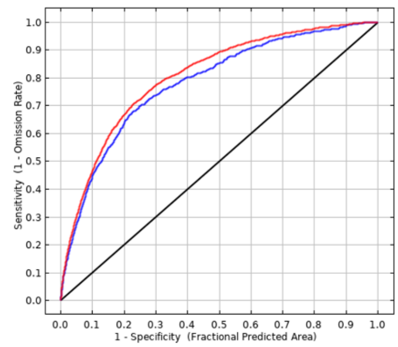

Figure 28 - Secondary Conductor ROC Curve

3.5.1.5.3 Model Data

Table 11 presents a summary of event counts used for v4 development:

Table 11 - Secondary Conductor Event Summary

Event Type	Value
Outage Count	3318
Ignition Count	223
Ignitions Per Outage	0.0672

3.5.1.5.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.1.5.5 Covariate Selection

Figure 29 presents the total gains feature importance that account for 95% of the Secondary Conductor model performance. Top features include strike tree counts, the impervious and non-burnable land indices that indicate urban environments, and strike tree height.

There are numerous additional covariate features that marginally contribute to the Other

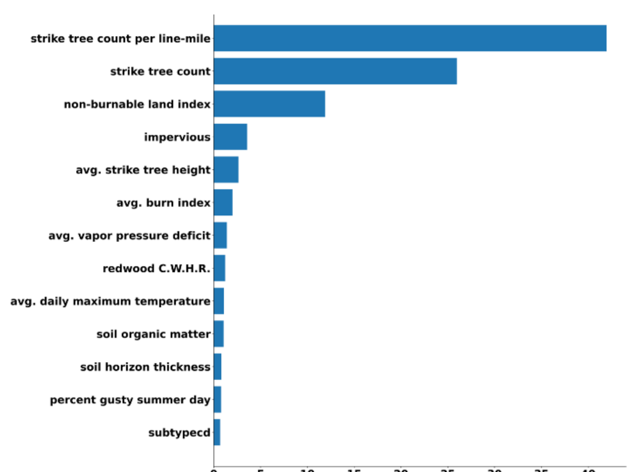

Figure 29 - Secondary Conductor Top Feature Performance

Equipment model, accounting for the remaining 5% of model performance:

- eucalyptus CWHR
- maximum strike tree height
- maximum wind speed
- avg. 100-hr fuels
- windy summer day percentage
- tier-2
- valley oak woodland CWHR
- minimum strike tree height
- montane hardwood CWHR
- soil slope
- available soil water capacity
- soil permeability
- coastal oak woodland CWHR
- annual soil flood frequency
- local topography
- canopy density
- tree species
- eastside pine CWHR
- ponderosa pine CWHR
- is soil hydric
- soil clay content
- operating voltage
- montane hardwood CWHR
- blue oak woodland CWHR
- avg. specific humidity
- soil hydro-group
- avg. precipitation
- aspen CWHR
- montane riparian CWHR
- closed cone pine cypress CWHR
- avg. wind speed
- avg. energy release

3.5.1.5.6 Limitations and Opportunities

Alternative deep learning algorithms such as Recurrent Neural Networks and Transformers are being investigated for future releases to improve the Secondary Conductor model with temporal dependencies in addition to the modeled spatial dependencies.

3.5.1.6 Support Structure Model

Support Structures physically support and maintain the alignment and positioning of distribution asset infrastructure. Support structure failures were modeled by asset for v3 using a time series algorithm, though the asset results were mapped to pixels for aggregation and compositing to circuit segments. The two causal failure models were refreshed for v4:

- Equipment Failure – any structural failure of a pole or its crossarms
- Electrical Failure – insulator or crossarm contamination, tracking, and other electrical issues

Similar to v3, the v4 scope of the Support Structure models is limited to wood poles only.

3.5.1.6.1 Support Structure – Electrical Failure Model

Support Structure electrical failures often result from contamination on crossarms and insulators that eventually leads to tracking and arcing particularly for older and cracked insulators. The failures frequently result in an ignition, but not necessarily an outage, making failures both a wildfire and a public safety concern. Fortunately, these ignitions are typically contained to the top of the pole and do not result in a fire large enough to be CPUC-reportable.

3.5.1.6.1.1 Key Developments

Key developments for the Support Structure – Electrical Failure model include:

- The electrical failure events training data set used for v3, which covered fire season events from 2016 through 2021, was updated to include the 2022 failure events for v4 model development.
- Seasonal meteorology attributes were developed as new model features for v4. Otherwise, the v4 model was mostly a refresh of the v3 model with updated failures and attributes data.
- Electrical failure probabilities and risks are estimated at the asset level for v4 and subsequently aggregated and composited through asset to conductor relationships to produce more accurate v4 circuit segment failure probability and risk values. While electrical failures were also modeled at an asset level for v3, the asset level results were converted into spatial pixel values for circuit segment aggregation and compositing. The v3 spatial aggregation frequently resulted in the electrical failure probability and risk for any given support structure being shared equally with every conductor asset that occupied the same pixel as the support structure. As a result, the real failure probability and risk values for multiple conductors passing through a single spatial pixel could be higher or lower than the reported aggregated values.

3.5.1.6.1.2 Performance

The new v4 Support Structure – Electrical failure model demonstrated very good model performance, as shown in *Figure 30* and *Figure 31*, with an AUC of 0.76.

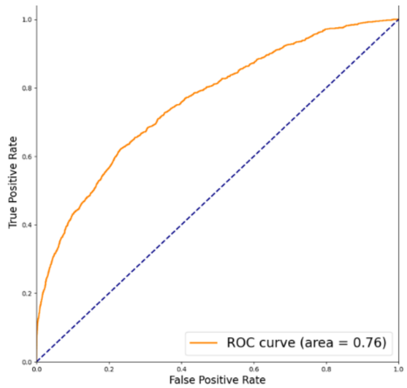

Figure 30 - Support Structure - Electrical ROC Curve

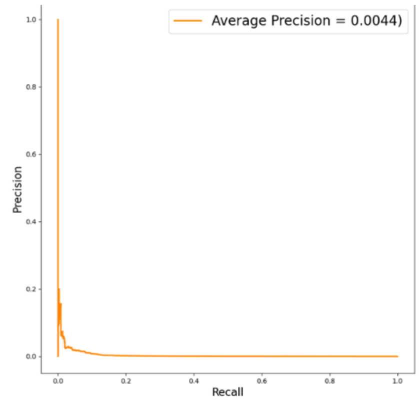

Figure 31 - Support Structure - Electrical Precision-Recall Curve

3.5.1.6.1.3 Model Data

The scope of the model was limited to wooden support structures on the distribution system. On average, there were 467 electrical support structure failures and 103 ignitions during wildfire season between 2016 – 2022.

Table 12 - Support Structure - Electrical Event Summary

Electrical Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	266	24	110	9
2017	306	19	108	8
2018	284	25	101	12
2019	241	18	91	10
2020	278	40	117	17
2021	632	272	270	109
2022	322	102	146	56
Totals	**2329**	**500**	**943**	**221**

3.5.1.6.1.4 Algorithm Selection

During development Support Structure – Electrical failure models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.6.1.5 Covariate Selection

Figure 32 presents the total gains top feature importance for the Support Structure – Electrical model performance. Top features include pole volume, age, humidity, non-burnable land index, precipitation, and tree count.

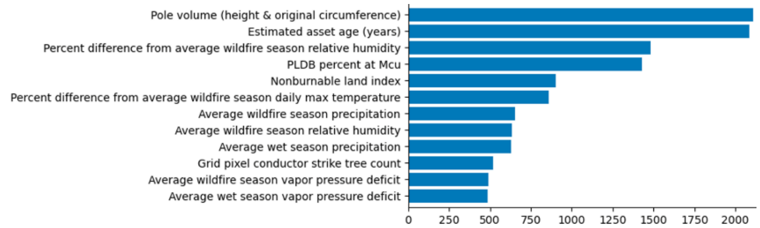

Figure 32 - Support Structure - Electrical Top Feature Importance

Other covariate features that moderately contribute to the Electrical failure model include:

- East Bay division
- Average wildfire season daily max windspeed
- Percent difference from average wet season relative humidity
- Percent difference from average wildfire season daily max windspeed
- Corrosion zone indicator
- Local topography
- Western Red Cedar pole species
- Percent difference from average wet season vapor pressure deficit
- Grid pixel resident count
- Soil annual flood frequency
- Average wet season relative humidity
- Average wildfire season daily max temperature
- Years since last PT&T inspection
- Mission division
- North Bay division
- Soil erodibility factor
- Percent difference from average wildfire season vapor pressure deficit
- Peninsula division
- Gas as original treatment type
- Sacramento division
- Unknown original treatment type
- San Francisco division
- Sierra division
- Not Available pole species
- Penta as original treatment type
- Central Coast division
- Pole attachment count

Figure 33 shows the SHAP plot for the Support Structure – Electrical failure model. Interpreting the plot, model prediction relationships can be understood and validated. For the support structures, higher probability of electrical failure was found to be correlated with:

- Low pole volumes
- High asset age
- Higher difference from average relative humidity

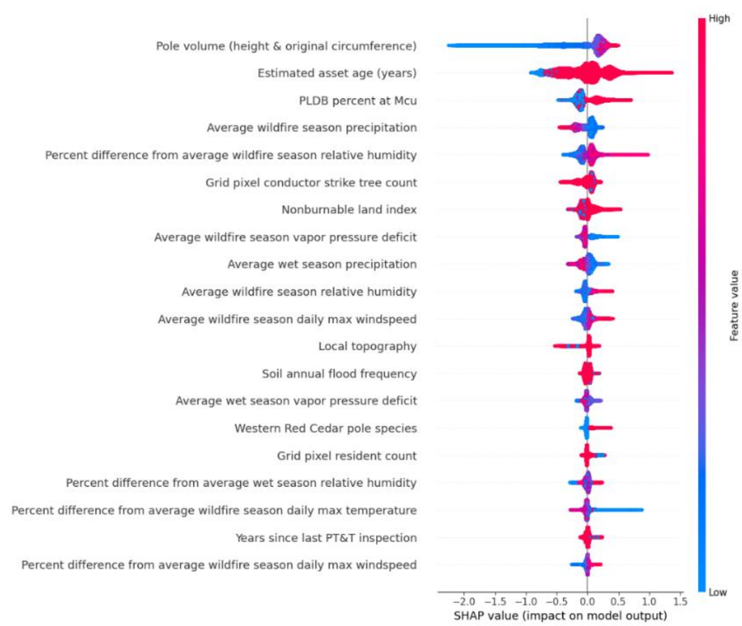

Figure 33- Support Structure - Electrical SHAP Plot

3.5.1.6.1.6 Limitations and Opportunities

As mentioned previously, the Support Structure model scope for v4 only considers wood poles. Work planners have frequently asked for the model coverage to be extended to include steel, concrete, and other types of poles as well.

Several initiatives could be pursued to improve the electrical failure model performance for future releases by exploring potential new model features, including:

- Large rain event identification – can wash away contamination
- Fog event identification – trigger tracking and arcing events
- Time-based air quality – predict contamination on insulators and crossarms
- Insulator type – identify insulators that are more prone to tracking and contamination

3.5.1.6.2 Support Structure – Structural Failure Model

Structural support structure failures occur when the support structure experiences a physical failure to the pole or its crossarm(s) and are often triggered by wind events. Structural failures have a low ignition to failure rate compared to other assets but often present a significant public safety issue. Typically, structural failures are the result of a pole ageing or decaying to the point where it can no longer carry its overhead equipment load. Decay is often accelerated when there is cycling between wet and dry soil in the subsurface portion of the pole.

3.5.1.6.2.1 Key Developments

Key developments for the Support Structure – Structural Failure model include:

- The structural failure events training data set used for v3, which covered fire season events from 2016 through 2021, was updated to include the 2022 failure events for v4 model development.
- Seasonal meteorology attributes were developed as new model features for v4. Otherwise, the v4 model was mostly a refresh of the v3 model with updated failures and attributes data.
- Structural failure probabilities and risks are estimated at the asset level for v4 and subsequently aggregated and composited through asset to conductor relationships to produce more accurate v4 circuit segment failure probability and risk values. While structural failures were also modeled at an asset level for v3, the asset level results were converted into spatial pixel values for circuit segment aggregation and compositing. The v3 spatial aggregation frequently resulted in the structural failure probability and risk for any given support structure being shared equally with every conductor asset that occupied the same pixel as the support structure. As a result, the real failure probability and risk values for multiple conductors passing through a single spatial pixel could be higher or lower than the reported aggregated values.

3.5.1.6.2.2 Performance

The new v4 Support Structure – Structural failure model demonstrated good model performance, as shown in *Figure 34* and *Figure 35*, with an AUC of 0.69.

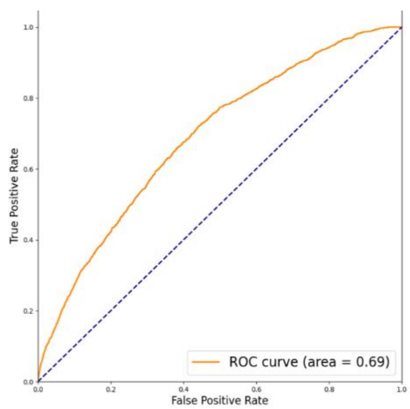

Figure 34 - Support Structure - Structural ROC Curve

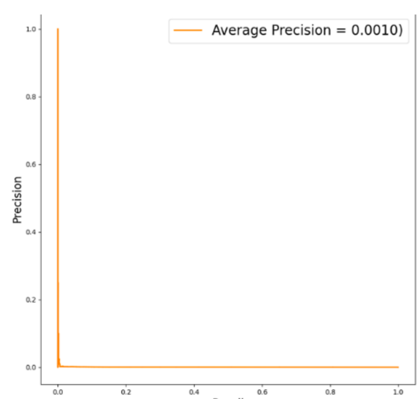

Figure 35 - Support Structure - Structural Precision-Recall Curve

3.5.1.6.2.3 Model Data

The scope of the model was limited to wooden support structures on the distribution system. On average, there were 667 structural support structure failures and 26.6 ignitions during wildfire season between 2016 – 2022.

Table 13 - Support Structure - Structural Event Summary

Structural Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	299	3	132	2
2017	365	5	169	2
2018	414	58	183	22
2019	518	11	203	2
2020	418	20	230	11
2021	625	33	238	8
2022	622	5	255	4
Totals	**3261**	**135**	**1410**	**51**

3.5.1.6.2.4 Algorithm Selection

During development Support Structure – Structural failure models were built using SparkML's Random Forest, Gradient Boosted Trees, Logistic Regression, and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.6.2.5 Covariate Selection

Figure 36 presents the total gains top feature importance for the Support Structure – Structural model performance. Top features include pole volume, pole age, pole strength, humidity, non-burnable land index, precipitation, and windspeed.

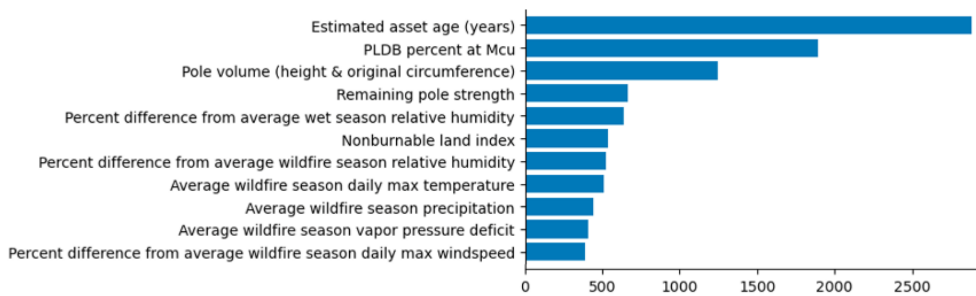

Figure 36 - Support Structure - Structural Top Feature Importance

Other covariate features that moderately contribute to the Structural failure model include:

- Average wildfire season relative humidity
- Difference between PT&T measured and original circumference
- Years since last PT&T inspection
- Percent difference from average wildfire season vapor pressure deficit
- Average wildfire season daily max windspeed
- Percent difference from average wildfire season daily max temperature
- Soil erodibility factor
- Average wet season relative humidity
- Average wet season precipitation
- Average wet season vapor pressure deficit
- Percent difference from average wet season vapor pressure deficit
- Unknown pole species
- Local topography * soil erodibility factor
- Open tag indicator
- Unknown original treatment type
- Grid pixel conductor strike tree count
- Soil annual flood frequency
- Local topography
- Penta as original treatment type
- Corrosion zone indicator

Figure 37 shows the SHAP plot for the Support Structure – Structural failure model. Interpreting the plot, model prediction relationships can be understood and validated. For the support structures, higher probability of structural failure was found to be correlated with:

- Low remaining pole strength
- Low percent at Mcu
- High asset age

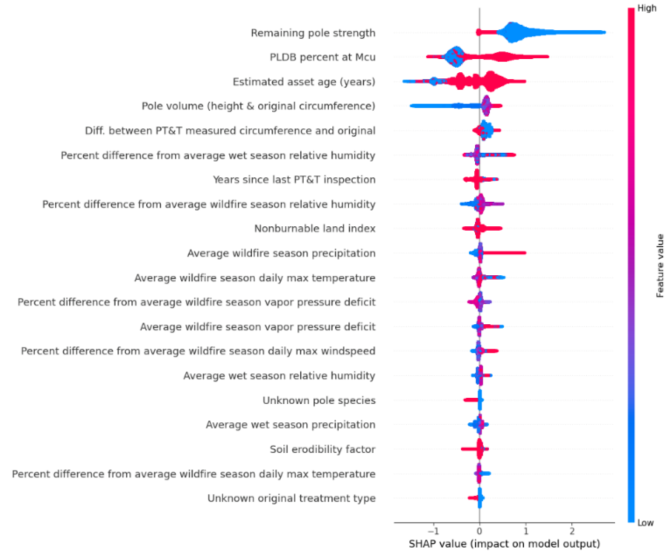

Figure 37 - Support Structure - Structural SHAP Plot

3.5.1.6.2.6 Limitations and Opportunities

As mentioned previously, the Support Structure model scope for v4 only considers wood poles. Work planners have frequently asked for the model coverage to be extended to include steel, concrete, and other types of poles as well.

Several initiatives could be pursued to improve the structural failure model performance for future releases by updating the model features, including:

- Additional data cleaning of the pole test and treat data set
- Testing removal of highly correlated features, especially meteorological features
- Additional data cleaning of structural failure events, with potential separation of:
    - Pole failure events
    - Crossarm failure events
    - Guywire failure events

3.5.1.7 Switch Equipment Model

Switches are devices that are used to control the flow of electricity in the network by opening or closing a section of a circuit. A switch can also serve to separate two different substation feeds. Switches are essential both for normal, emergency, and maintenance operations of the distribution network. Switches can fail from the loosening of hardware, causing operational problems of the switch, insulator breakage, or failed connectors. Switches have a relatively high ignition to failure rate when compared with other distribution assets.

3.5.1.7.1 Key Developments

Fuse, DPD equipment, primary conductor, and switch events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, Switches were modeled separately from fuses, conductors, and DPDs. Additionally, the v3 spatial model approach was replaced with a time-series model for v4.

The time-series model upgrade was enabled by establishing switch asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, failures could be traced back to specific equipment in service at the time of failure and the actual attributes of the failed equipment could in turn be used to train the model.

During model development there were some features that were explored, but not included as covariates in the final model:

- Open tags
- Seasonal meteorology data

3.5.1.7.2 Performance

The v4 Switch failure model demonstrated a moderately good model performance, as shown in *Figure 38* and *Figure 39*, with an AUC of 0.67.

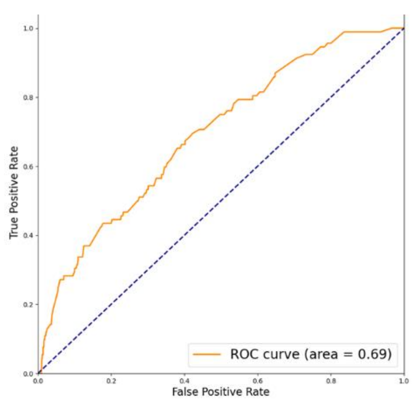

Figure 38 - Switch ROC Curve

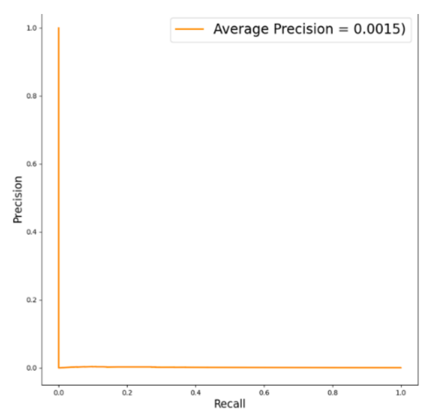

Figure 39 - Switch Precision-Recall Curve

3.5.1.7.3 Model Data

The scope of the model was limited to overhead switches on the distribution system. On average, there were 40 switch failures and 5 ignitions during wildfire season between 2016 – 2022.

Table 14 - Switch Equipment Event Summary

Switch Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	32	7	14	1
2017	28	4	15	2
2018	21	0	14	0
2019	18	3	13	1
2020	25	3	14	1
2021	37	6	13	0
2022	28	6	9	1
Totals	**189**	**29**	**92**	**6**

3.5.1.7.4 Algorithm Selection

During development Switch failure models were built using SparkML's Random Forest and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.7.5 Covariate Selection

Many switch equipment attributes were tested as features in the model, with the SCADA switch type showing the most performance gains. *Figure 40* presents the total gains feature importance for the Switch equipment model. Among the many equipment attributes experimented with during model development: switch CC rating, soil erodibility, average fire season humidity, and disconnect switch type were found to be the most meaningful features for the model.

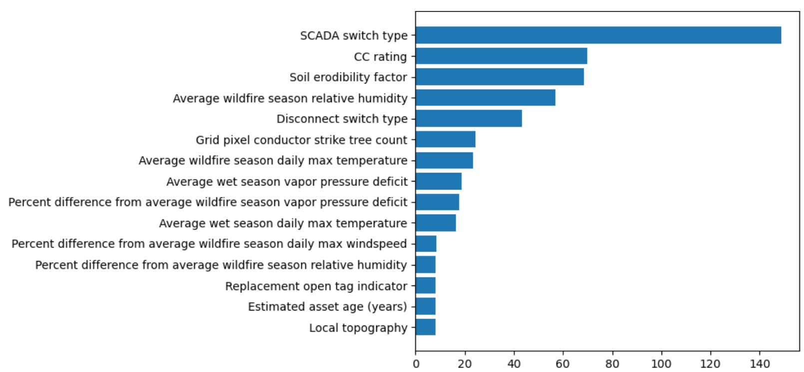

Figure 40 - Switch Feature Importance

Figure 41 shows the SHAP plot for the Switch model. Interpreting the plot, model prediction relationships can be understood and validated. For the Switch equipment, higher probability of failure was found to be correlated with:

- Higher CC rating
- Not disconnect switch
- Is SCADA switch

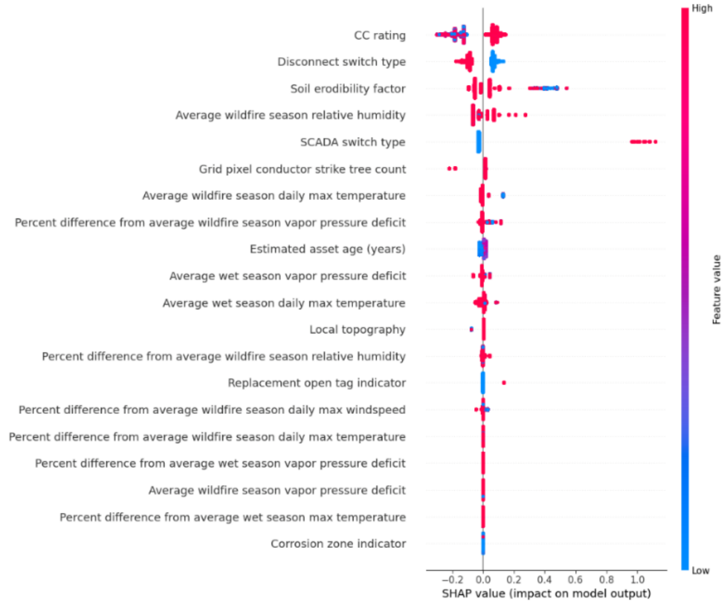

Figure 41 - Switch SHAP Plot

3.5.1.7.6 Limitations and Opportunities

The new v4 switch model demonstrated the capability to use switch attributes with environmental and meteorological features to predict switch equipment failures. Future model development will explore potential improvements such as:

- Using the device switching history of the device - a healthy switch is operated only occasionally and switches that operate frequently tend to degrade.
- Additional research on why SCADA switches are more likely to fail.
- Improving CC rating data set quality.

3.5.1.8 Transformer Models

Transformers are nodes on the distribution grid that are identified by a California Grid Coordinate (cgc12). Each transformer contains from one to three transformer units that are identified by individual equipment IDs. When any unit on a transformer fails, all individual transformer units are typically replaced. Planning programs, therefore, focus on the transformer, whereas failures occur to a transformer unit. The transformer models therefore predict failures for individual transformer units.

The transformer failures are predicted through two causal models for v4:

- Transformer – Equipment failures caused by overload or equipment issues
- Transformer – Leaking failures caused by oil escaping from the transformer unit

3.5.1.8.1 Transformer – Equipment Model

The Transformer – Equipment model predicts physical transformer equipment failures such as winding failures, core failures and loss of neutral. Transformers can potentially fail catastrophically with an explosion that is both a wildfire and a public safety concern. Fortunately, transformers have a very low wildfire ignition to failure rate when compared with other distribution assets.

3.5.1.8.1.1 Key Developments

Key developments for the Transformer – Equipment Failure model include:

- The equipment failure events training data set used for v3, which covered fire season events from 2019 through 2021, was updated to include the 2022 failure events for v4 model development.
- New attributes for v4:
  - Replaced v3 categorized loading values with historical loading data.
  - Open transformer tags
  - Seasonal meteorology
- Equipment failure probabilities and risks are estimated at the asset level for v4 and subsequently aggregated and composited through asset to conductor relationships to produce more accurate v4 circuit segment failure probability and risk values. While structural failures were also modeled at an asset level for v3, the asset level results were converted into spatial pixel values for circuit segment aggregation and compositing. The v3 spatial aggregation frequently resulted in the structural failure probability and risk for any given support structure being shared equally with every conductor asset that occupied the same pixel as the support structure. As a result, the real failure probability and risk values for multiple conductors passing through a single spatial pixel could be higher or lower than the reported aggregated values.

3.5.1.8.1.2 Performance

The v4 Transformer - Equipment failure model demonstrated very good model performance, as shown in *Figure 42* and *Figure 43*, with an AUC of 0.77.

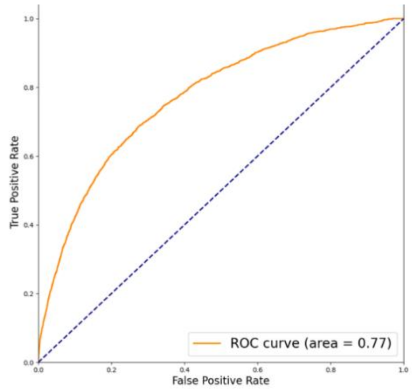

Figure 42 - Transformer - Equipment ROC Curve

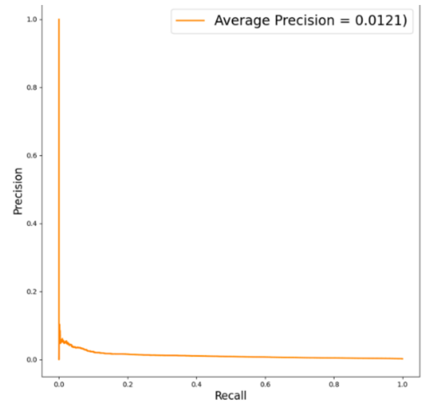

Figure 43 - Transformer - Equipment Precision-Recall Curve

3.5.1.8.1.3 Model Data

The scope of the model was limited to overhead transformer equipment on the distribution system. On average, there were 1,960 transformer equipment failures and 20.5 ignitions during wildfire season between 2019 – 2022.

Table 15 - Transformer - Equipment Event Summary

Transformer - Equipment Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2019	1097	18	391	7
2020	1662	18	545	2
2021	1603	15	511	6
2022	1533	14	498	2
Totals	**5895**	**65**	**1945**	**17**

3.5.1.8.1.4 Algorithm Selection

During development Switch failure models were built using SparkML's Random Forest and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.8.1.5 Covariate Selection

Figure 44 presents the total gains top feature importance for the Transformer – Equipment model performance. Top features include peak kVA, transformer age, temperature, resident count, and load factor.

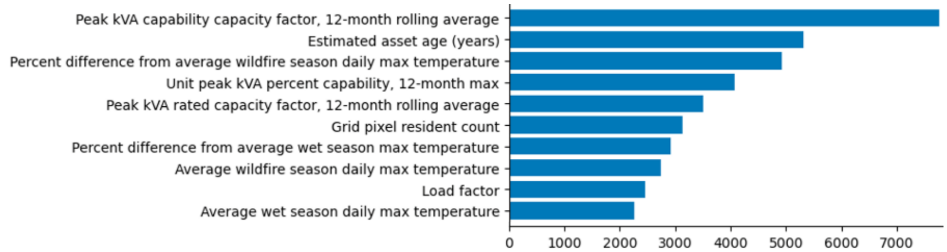

Figure 44 - Transformer - Equipment Top Feature Importance

Other covariate features that moderately contribute to the Equipment failure model include:

- Average wildfire season daily max windspeed
- Soil erodibility factor
- Grid pixel conductor strike tree count
- Local topography
- Rated kVA
- Unit peak kVA percent capability, 12-month rolling average
- Coastal corrosive indicator
- Non-burnable land index
- Manufactured encoded KU
- Manufactured encoded AS
- Manufactured encoded PP
- Manufactured encoded GE
- Manufactured encoded HI
- Manufactured encoded DS
- Manufactured encoded XX
- Replacement open tag indicator
- Manufactured encoded ME
- Manufactured encoded FP
- Manufactured encoded CP
- Manufactured encoded AB
- Manufactured encoded WE

Figure 45 shows the SHAP plot for the Transformer – Equipment failure model. Interpreting the plot, model prediction relationships can be understood and validated. For transformer equipment, higher probability of failure was found to be correlated with:

- Higher peak kVA rated capacity factor
- High percent difference from average max wet season temperature

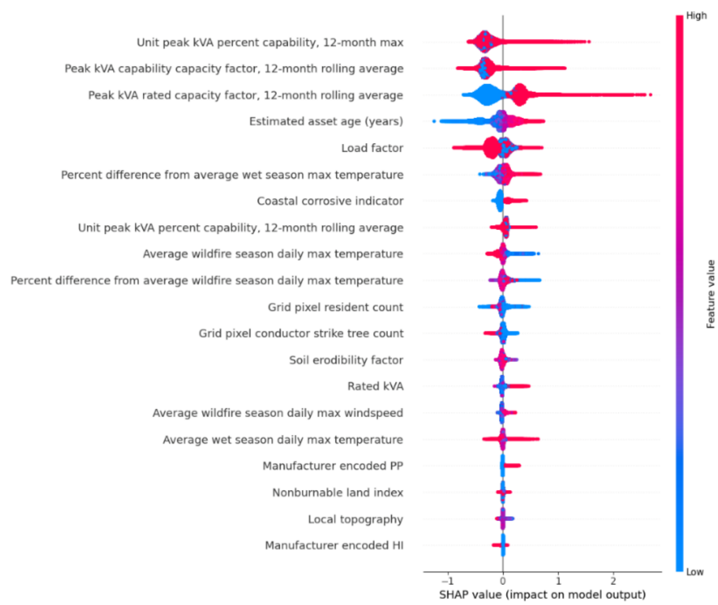

Figure 45 - Transformer - Equipment SHAP Plot

3.5.1.8.1.6 Limitations and Opportunities

Several initiatives could be pursued to improve the transformer equipment failure model performance for future releases by updating the model features, including:

- Additional cleaning of failure and ignition event data to identify specific unit failures
- Explore incorporating additional historical loading data
- Explore incorporating physics-based models for temperature-based adjustments to asset age

3.5.1.8.2 Transformer – Leaking Model

The Transformer – Leaking model predicts tank failures that result in cooling oil leaking out of a transformer. Cooling oil loss leads to overheated equipment and eventual failure of the transformer. While transformer leak failures rarely lead to ignitions, modeling leak failures separately improves the performance of the equipment failure model for predicting equipment-related ignitions.

3.5.1.8.2.1 Key Developments

Key developments for the Transformer – Leaking model include:

- The equipment leaking events training data set used for v3, which covered fire season events from 2019 through 2021, was updated to include the 2022 failure events for v4 model development.
- New attributes for v4:
  - Replaced v3 categorized loading values with historical loading data.
  - Open transformer tags
  - Seasonal meteorology
- Equipment failure probabilities and risks are estimated at the asset level for v4 and subsequently aggregated and composited through asset to conductor relationships to produce more accurate v4 circuit segment failure probability and risk values. While structural failures were also modeled at an asset level for v3, the asset level results were converted into spatial pixel values for circuit segment aggregation and compositing. The v3 spatial aggregation frequently resulted in the structural failure probability and risk for any given support structure being shared equally with every conductor asset that occupied the same pixel as the support structure. As a result, the real failure probability and risk values for multiple conductors passing through a single spatial pixel could be higher or lower than the reported aggregated values.

3.5.1.8.2.2 Performance

The v4 Transformer - Leaking failure model demonstrated very good model performance, as shown in *Figure 46* and, *Figure 47* with an AUC of 0.78.

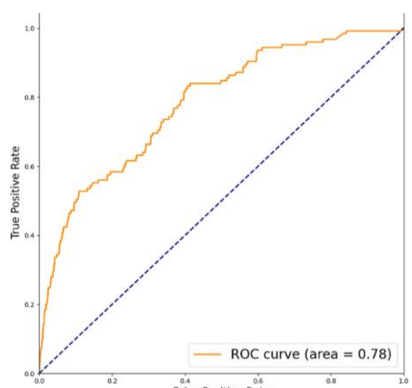

Figure 46 - Transformer - Leaking ROC Curve

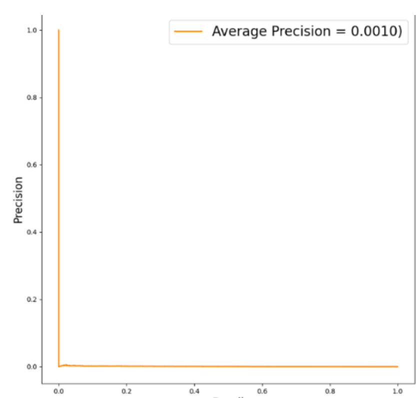

Figure 47 - Transformer - Leaking Precision-Recall Curve

3.5.1.8.2.3 Model Data

The scope of the model was limited to overhead transformer equipment on the distribution system. On average, there were 113 transformer leaking failures and less than one ignition during wildfire season between 2019 – 2022.

Table 16 - Transformer - Leaking Event Summary

Transformer - Leaking Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2019	49	0	21	0
2020	88	0	32	0
2021	75	1	22	0
2022	114	0	50	0
Totals	**326**	**1**	**125**	**0**

3.5.1.8.2.4 Algorithm Selection

During development Switch failure models were built using SparkML's Random Forest and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.8.2.5 Covariate Selection

Figure 48 presents the total gains top feature importance for the Transformer – Leaking model performance. Top features include temperature, transformer age, peak KVA, rate kVA resident count, load factor, windspeed, and soil erodibility factor.

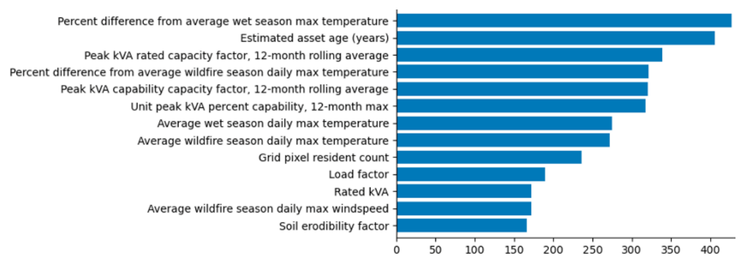

Figure 48 - Transformer - Leaking Top Feature Importance

Other covariate features that moderately contribute to the Leaking failure model include:

- Manufacturer encoded DS
- Local topography
- Grid pixel conductor strike tree count
- Unit peak kVA percent capability, 12-month rolling average
- Manufactured encoded HI
- Manufactured encoded XX
- Manufactured encoded AS
- Manufactured encoded PP
- Non-burnable land index
- Coastal corrosive indicator
- Manufactured encoded GE
- Manufactured encoded KU
- Manufactured encoded ME

Figure 49 shows the SHAP plot for the Transformer – Leaking failure model. Interpreting the plot, model prediction relationships can be understood and validated. For leaking transformers, higher probability of failure was found to be correlated with:

- Higher peak kVA rated capability capacity factor
- Higher peak kVA rated capacity factor
- Unit peak kVA percent capability
- Transformer age

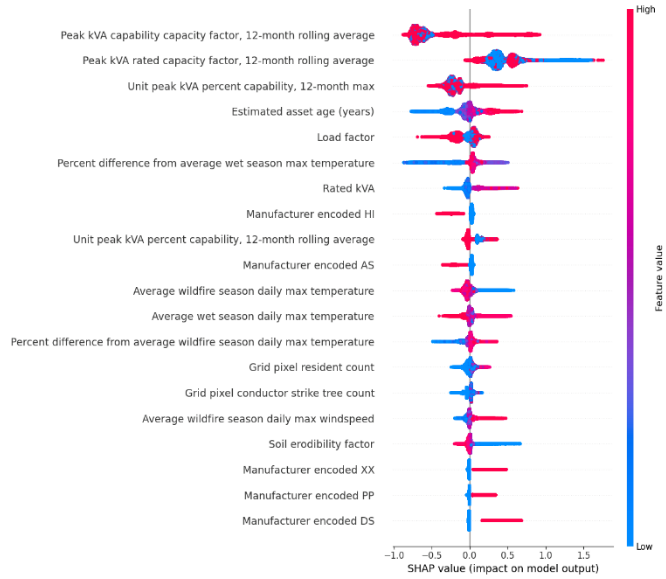

Figure 49 - Transformer - Leaking SHAP Plot

3.5.1.8.2.6 Limitations and Opportunities

Several initiatives could be pursued to improve the transformer equipment failure model performance for future releases by updating the model features, including:

- Additional cleaning of failure and ignition event data to identify specific unit failures
- Explore incorporating additional historical loading data
- Explore incorporating physics-based models for temperature-based adjustments to asset age

3.5.1.9 Voltage Regulator Model

Voltage regulators, also called boosters, help keep the voltage steady on the distribution system. They have a moderate ignition to failure rate compared to other distribution assets.

3.5.1.9.1 Key Developments

Voltage regulator and capacitor bank failure events were bundled together in a common spatial model for v3 using the MaxEnt algorithm. For v4, voltage regulator and capacitor bank failure events are modeled separately. Additionally, the v3 spatial model approach was replaced with a time-series model for v4.

The time-series model upgrade was enabled by establishing voltage regulator asset history back to 2016 through an examination of annual snapshots of Electrical Distribution GIS (EDGIS) asset data. By examining the historical asset data, failures could be traced back to specific equipment in service at the time of failure and the actual attributes of the failed equipment could in turn be used to train the model.

During model development there were some features that were explored, but not included as covariates in the final model:

- Open tags
- Seasonal meteorology data

3.5.1.9.2 Performance

The v4 Voltage Regulator failure model demonstrated moderately good model performance, as shown in *Figure 50* and *Figure 51*, with an AUC of 0.62.

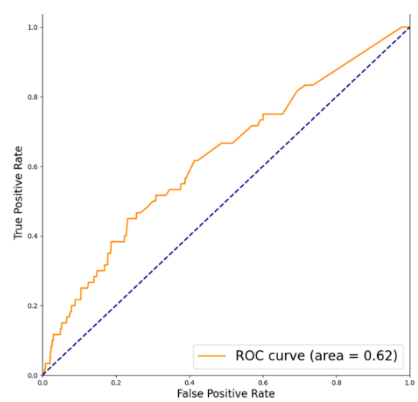

Figure 50 - Voltage Regulator ROC Curve

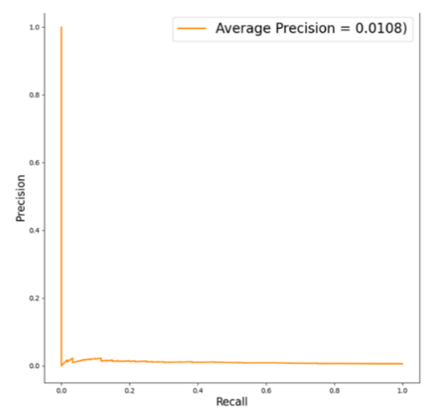

Figure 51 - Voltage Regulator Precision-Recall Curve

3.5.1.9.3 Model Data

The scope of the model was limited to overhead voltage regulators on the distribution system. On average, there were 30.5 voltage regulator failures and 2.7 ignitions during wildfire season between 2016 – 2022.

Table 17 - Voltage Regulator Event Summary

Voltage Regulator Event Counts				
	Training Data		Testing Data	
Year	Failures	Ignitions	Failures	Ignitions
2016	17	1	6	0
2017	19	1	14	2
2018	13	3	7	0
2019	26	1	6	0
2020	21	4	9	1
2021	33	3	8	1
2022	25	1	10	1
Totals	**154**	**14**	**60**	**5**

3.5.1.9.4 Algorithm Selection

During development Voltage Regulator failure models were built using SparkML's Random Forest and the Spark version of XGBoost. Ultimately, the XGBoost algorithm was selected for its predictive performance, built-in feature selection, ability to handle missing values, and speedy computation time.

3.5.1.9.5 Covariate Selection

Many voltage regulator attributes were tested as features in the model with the rate amps showing the most performance gains. *Figure 52* presents the total gains feature importance for the Voltage Regulator model. Among the many equipment attributes experimented with during model development: wet season precipitation, wet season maximum daily temperature, wet season vapor pressure deficit, wildfire season humidity, and percent difference from wildfire season humidity were found to be the most meaningful features for the model.

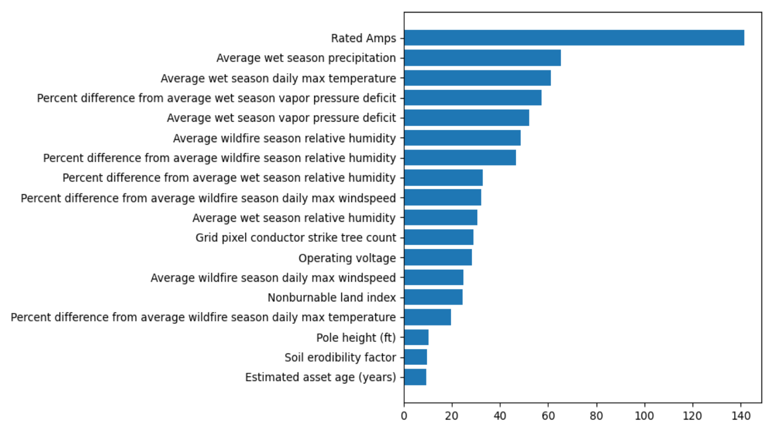

Figure 52 - Voltage Regulator Feature Importance

Figure 53 shows the SHAP plot for the Voltage Regulator model. Interpreting the plot, model prediction relationships can be understood and validated. For Voltage Regulators, higher probability of failure was found to be correlated with:

- Lower wet season precipitation
- Lower fire season humidity
- Higher rated amp values

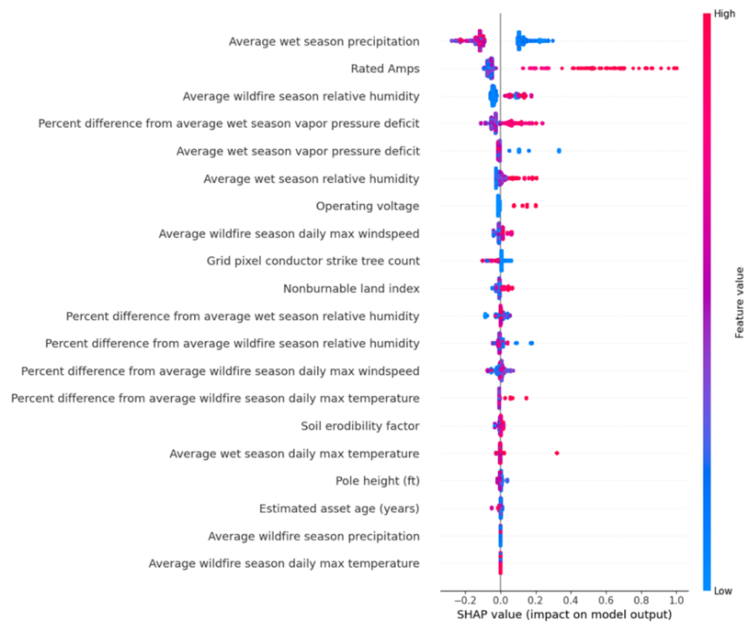

Figure 53 - Voltage Regulator SHAP Plot

3.5.1.9.6 Limitations and Opportunities

The Voltage Regulator model is limited to banks, or groups, of voltage regulators. Unit-specific failure modeling would require additional data not currently available.

Future model development is expected to focus on the exploration of additional voltage regulator attributes and meteorological features.

3.5.1.10 Other Equipment Model

The Other Equipment model is the bookkeeping catch-all for outages with a cause that does not belong with any other causal model or has no recorded cause. Fortunately, while there are a great number of outages that are attributed to the Other Equipment model, there are only a very few outages that result in an ignition. This is likely due to the investigative emphasis placed on ignition records leading to correct attribution of each ignition to its true cause. In the end, the small number of ignitions assigned to the Other Equipment model have a minimal impact on distribution wildfire risk mitigation planning.

3.5.1.10.1 Key Developments

The Other Equipment v4 model is a refresh of the v3 model. The only significant update was the addition of 2023 failures to the event training dataset.

3.5.1.10.2 Performance

The v4 Other Equipment outage model demonstrated good model performance with an AUC of 0.75 as shown in *Figure 54*.

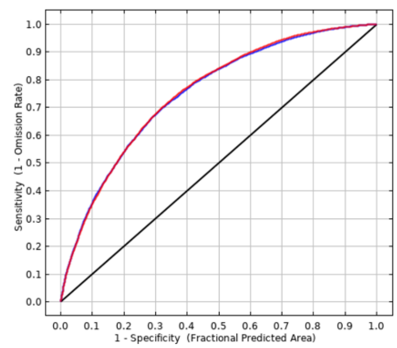

Figure 54 - Other Equipment ROC Curve

3.5.1.10.3 Model Data

Table 18 presents a summary of event counts used for v4 development:

Table 18 – Other Equipment Event Summary

Event Type	Value
Outage Count	54053
Ignition Count	201
Ignitions Per Outage	0.0037

3.5.1.10.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.1.10.5 Covariate Selection

Figure 55 presents the total gains feature importance that account for 95% of the Other Equipment model performance. Top features include strike tree counts, the impervious and non-burnable land indices that indicate urban environments, subtypecd, and conductor size and material.

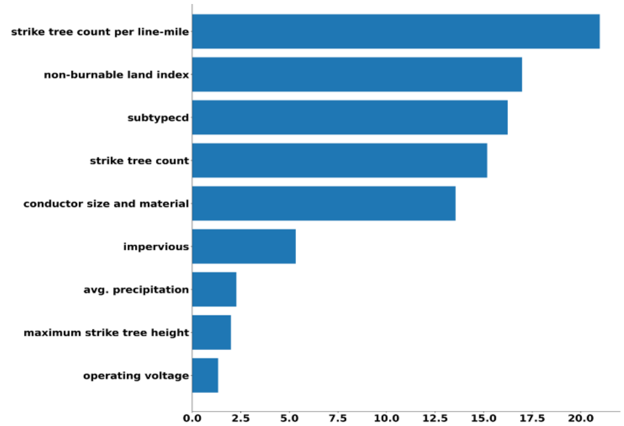

Figure 55 - Other Equipment Top Feature Performance

There are numerous additional covariate features that marginally contribute to the Other Equipment model, accounting for the remaining 5% of model performance:

- avg. strike tree height
- canopy density
- montane hardwood CWHR
- tier-2
- coastal
- soil slope
- P.S.P.S. segment
- montane hardwood CWHR
- soil clay content
- douglas fir CWHR
- soil erodibility factor
- avg. specific humidity
- gusty summer day percentage
- closed cone pine cypress CWHR
- soil drain
- soil horizon thickness
- soil hydro-group
- is soil hydric
- annual soil flood frequency
- windy summer day percentage
- redwood CWHR
- local topography
- soil permeability
- eucalyptus C.W.H.R
- klamath mixed conifer CWHR
- avg. 100-hr fuels
- tree species
- coastal oak woodland CWHR
- valley oak woodland CWHR
- juniper CWHR
- ponderosa pine CWHR
- montane riparian CWHR
- avg. burn index
- avg. daily maximum temperature
- maximum wind speed
- soil organic matter
- sierran mixed conifer CWHR

3.5.1.10.6 Limitations and Opportunities

Deep learning models will be investigated to bring these models into the temporal domain while retaining spatial dependencies.

3.5.2 Contact From Object Models

Contact From Object failures and outages on the distribution grid occur when an electrical short or failure is caused by a non-electrical asset object. Contact From Object failures are grouped into three causal categories:

- Animal - Failures initiated by animal (non-human) contact with an electrical asset.
- Third-Party - Failures initiated by human-related contact with an electrical asset.
- Vegetation - Failures initiated by vegetation in contact with an electrical asset.

Thousands of failures and outages have been recorded for each causal category. To improve model performance and model applicability for developing mitigation response, each of the causal categories have been subdivided to model specific causal failure paths.

The following sections in this document detail the suite of Contact From Object models developed for DEPM v4.

Table 19 provides a brief Contact From Object model performance summary.

Table 19 - Contact From Object Model Performance Summary

Contact From Object Event Model	Outages	v3 p(o)	V4 p(o) – Probability of Outage		
	Season Mean	AUC	AUC	Precision	Concentration Factor
Vegetation – Branch		0.83	0.87	0.000864	3.20
Vegetation – Trunk		0.87	0.88	0.000971	3.40
Vegetation – Other		0.84	0.88	0.000357	2.57
Animal – Bird		0.74	0.75	0.000266	1.79
Animal – Squirrel		0.88	0.88	0.000174	3.70
Animal – Other		0.77	0.80	0.000195	2.53
Third-Party – Balloon		0.89	0.89	0.000305	3.25
Third-Party – Vehicle		0.77	0.76	0.000413	1.85
Third-Party – Other		0.71	0.74	0.000145	1.78

3.5.2.1 Animal Models

Animals frequently come into contact with electrical assets. Animal contact can cause outages and potentially ignitions. Historically, birds and squirrels are responsible for the majority of animal-related outages. Therefore, animal outages are predicted using three distinct models:

- Bird – Failures caused by avian contact with conductors.
- Squirrel – Failures caused by squirrel contact with conductors.
- Other – Failures from other or unknown animal contact.

3.5.2.1.1 Animal – Bird Model

The Animal-Bird model uses historical bird-caused failure events to model outage probability from bird contact with conductors.

3.5.2.1.1.1 Key Developments

There are two important changes to the Bird model for v4:

- The failure event dataset was updated to include 2023 failures, expanding the number of events available for model development.
- New satellite canopy density data from Planet Labs was introduced to improve location correlations for animal habitats.

3.5.2.1.1.2 Performance

The v4 Animal – Bird outage model demonstrated good model performance with an AUC of 0.75, as shown in *Figure 56*.

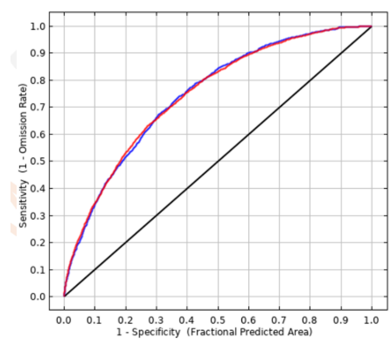

Figure 56 - Animal-Bird ROC Curve

3.5.2.1.1.3 Model Data

Table 20 presents a summary of event counts used for v4 development:

Table 20 - Animal - Bird Event Summary

Event Type	Value
Outage Count	5429
Ignition Count	259
Ignitions Per Outage	0.0477

3.5.2.1.1.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.1.1.5 Covariate Selection

Figure 57 presents the total gains feature importance that account for 95% of the Animal - Bird model performance. Top features include conductor type (subtypecd), the number of strike trees, impervious and non-burnable land index that indicate urban environments, and operating voltage.

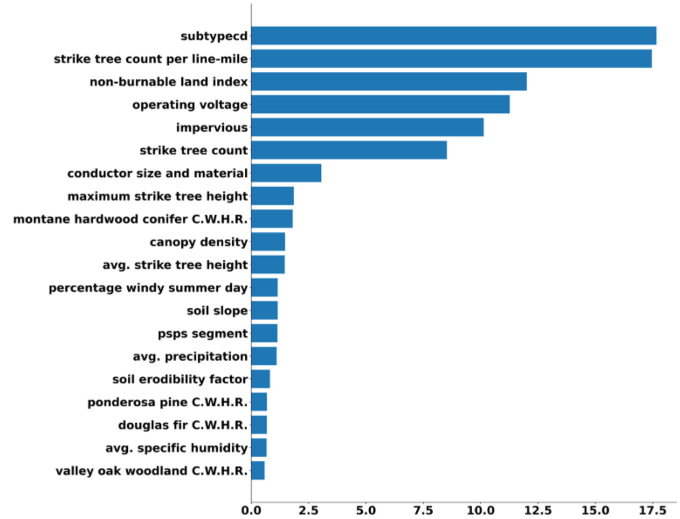

Figure 57 - Animal - Bird Top Feature Importance

There are numerous additional covariate features that marginally contribute to the Bird model, accounting for the remaining 5% of model performance:

- soil organic matter
- closed cone pine cypress CWHR
- tier-2
- tree species
- avg. vapor pressure deficit
- soil hydro-group
- blue oak foothill pine CWHR
- blue oak woodland CWHR
- local topography
- avg. 100-hr fuels
- soil clay content
- avg. 1000-hr fuels
- avg. energy release
- soil liquid limit
- gusty summer day percentage
- avg. wind speed
- minimum strike tree height
- tier-3
- soil permeability
- avg. daily maximum temperature
- soil drain
- soil horizon thickness
- eucalyptus C.W.H.R
- aspen C.W.H.R
- is soil hydric
- juniper CWHR
- available soil water capacity
- maximum wind speed
- montane hardwood CWHR
- annual soil flood frequency
- redwood CWHR

3.5.2.1.1.6 Limitations and Opportunities

Due to the interaction of animals, vegetation, and climate, it is important that the Animal models be able to capture temporal variations over time so that future conditions can be projected based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends. Therefore, the model is currently limited to using features such as overall vegetation density and cannot discern time-based trends such as increasing tree mortality and its resultant animal habitat impacts. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.1.2 Animal – Squirrel Model

The Animal-Squirrel model uses historical squirrel-caused failure events to model outage probability from squirrels interacting with conductors.

3.5.2.1.2.1 Key Developments

There are two important changes to the Squirrel model for v4:

- The failure event dataset was updated to include 2023 failures, expanding the number of events available for model development.
- New satellite canopy density data from Planet Labs was introduced to improve location correlations for animal habitats.

3.5.2.1.2.2 Performance

The v4 Animal – Squirrel outage model demonstrated excellent model performance with an AUC of 0.88, as shown in *Figure 58*.

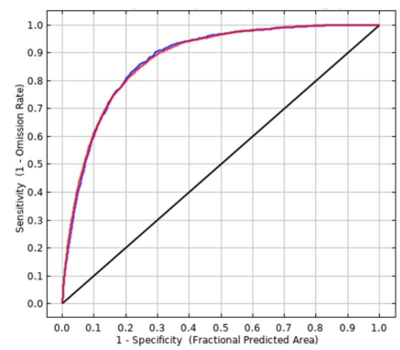

Figure 58 - Animal-Squirrel ROC Curve

3.5.2.1.2.3 Model Data

Table 21 presents a summary of event counts used for v4 development:

Table 21 - Animal-Squirrel Event Summary

Event Type	Value
Outage Count	4488
Ignition Count	51
Ignitions Per Outage	0.0114

3.5.2.1.2.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.1.2.5 Covariate Selection

Figure 59 presents the total gains feature importance that account for 95% of the Animal - Squirrel model performance. Top features include impervious and non-burnable land indices that indicate urban environments, average precipitation, strike tree count, and fuel availability.

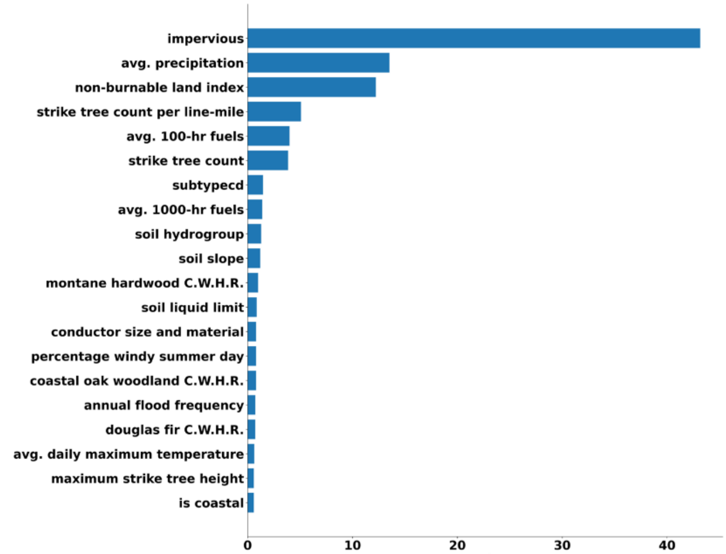

Figure 59 - Animal - Squirrel Top Feature Importance

There are numerous additional covariate features that marginally contribute to the Squirrel model, accounting for the remaining 5% of model performance:

- avg. specific humidity
- soil horizon thickness
- avg. vapor pressure deficit
- valley oak woodland CWHR
- blue oak woodland CWHR
- canopy density
- maximum wind speed
- soil erodibility factor
- eucalyptus C.W.H.R
- avg. strike tree height
- operating voltage
- montane hardwood CWHR
- P.S.P.S. segment
- tier-2
- montane riparian CWHR
- soil clay content
- gusty summer day percentage
- avg. wind speed
- sierran mixed conifer CWHR
- avg. burn index
- tree species
- valley foothill riparian CWHR
- local topography
- eastside pine CWHR
- closed cone pine cypress CWHR
- ponderosa pine CWHR
- soil permeability
- available soil water capacity
- is soil hydric

3.5.2.1.2.6 Limitations and Opportunities

Due to the interaction of animals, vegetation, and climate, it is important that the Animal models be able to capture temporal variations over time so that future conditions can be projected based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends. Therefore, the model is currently limited to using features such as overall vegetation density and cannot discern time-based trends such as increasing tree mortality and its resultant animal habitat impacts. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.1.3 Animal – Other Model

The Animal – Other model uses historical failure events which do not fit within the squirrel or bird category to model outage probability from other animal sources.

3.5.2.1.3.1 Key Developments

There are two important changes to the Animal – Other model for v4:

- The failure event dataset was updated to include 2023 failures, expanding the number of events available for model development.
- New satellite canopy density data from Planet Labs was introduced to improve location correlations for animal habitats.

3.5.2.1.3.2 Performance

The v4 Animal – Other outage model demonstrated very good model performance with an AUC of 0.79, as shown in *Figure 60*.

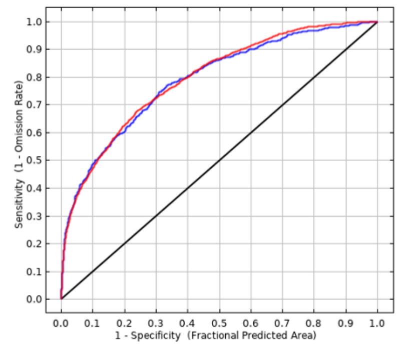

Figure 60 - Animal-Other ROC Curve

3.5.2.1.3.3 Model Data

Table 22 presents a summary of event counts used for v4 development:

Table 22 - Animal - Other Event Summary

Event Type	Value
Outage Count	1291
Ignition Count	105
Ignitions Per Outage	0.0813

3.5.2.1.3.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.1.3.5 Covariate Selection

Figure 61 presents the total gains feature importance that account for 95% of the Animal - Squirrel model performance. Top features include impervious and non-burnable land indices that indicate urban environments, strike tree count, vapor pressure deficit and humidity.

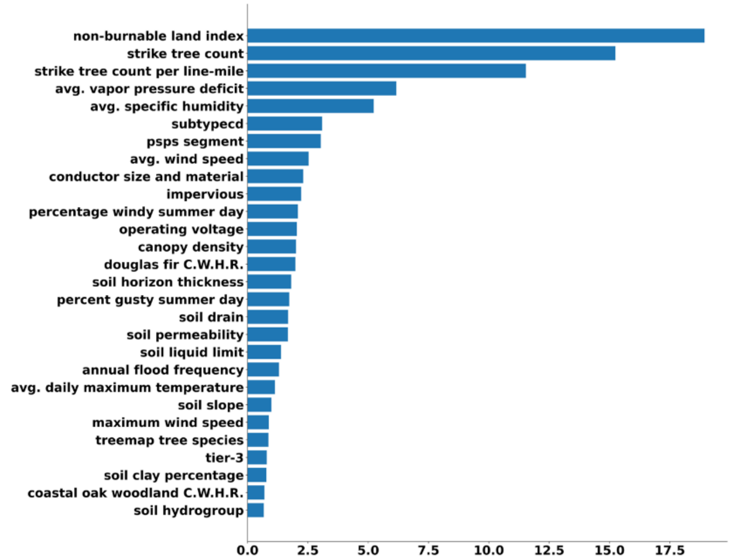

Figure 61 - Animal - Other Top Feature Performance

There are numerous additional covariate features that marginally contribute to the Other model, accounting for the remaining 5% of model performance:

- soil organic matter
- closed cone pine cypress CWHR
- soil erodibility factor
- blue oak woodland CWHR
- avg. burn index
- coastal
- avg. 1000-hr fuels
- available soil water capacity
- avg. strike tree height
- montane hardwood CWHR
- minimum strike tree height
- avg. precipitation
- avg. energy release
- eucalyptus C.W.H.R
- montane riparian CWHR
- valley oak woodland CWHR
- redwood CWHR
- is soil hydric

3.5.2.1.3.6 Limitations and Opportunities

Due to the interaction of animals, vegetation, and climate, it is important that the Animal models be able to capture temporal variations over time so that future conditions can be projected based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends. Therefore, the model is currently limited to using features such as overall vegetation density and cannot discern time-based trends such as increasing tree mortality and its resultant animal habitat impacts. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.2 Third-Party Models

Third-party failures, outages, and ignitions occur when there is human-related contact with electrical assets. Historically, vehicles and mylar balloons are responsible for the majority of human-related outages. Therefore, Third-Party outages are predicted using three distinct models:

- Balloon – Failures caused by mylar balloon contact with conductors.
- Vehicle – Failures caused by a vehicle contact with poles and conductors.
- Other – Failures from other or unknown human-related contact.

3.5.2.2.1 Third-Party – Balloon Model

The Third-Party Balloon model uses historical mylar balloon-caused failure events to model outage probability from balloons contacting conductors.

3.5.2.2.1.1 Key Developments

The only significant change to the Balloon model for v4 was that the failure event data set was updated to include 2023 failures, increasing the number of events used to for modeling training.

3.5.2.2.1.2 Performance

The v4 Third-Party – Balloon outage model demonstrated excellent model performance with an AUC of 0.88, as shown in *Figure 62*.

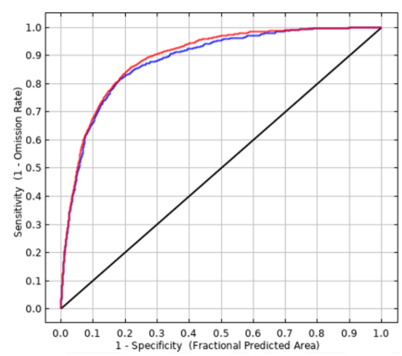

Figure 62 - Third-Party – Balloon ROC Curve

3.5.2.2.1.3 Model Data

Table 23 presents a summary of event counts used for v4 development:

Table 23 - Third-Party – Balloon Event Summary

Event Type	Value
Outage Count	2420
Ignition Count	120
Ignitions Per Outage	0.0496

3.5.2.2.1.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.2.1.5 Covariate Selection

Figure 63 presents the total gains feature importance that account for 95% of the Third-Party - Balloon model performance. Top features include the impervious and non-burnable land indices that indicate urban environments, strike tree count, and conductor type (subtypecd).

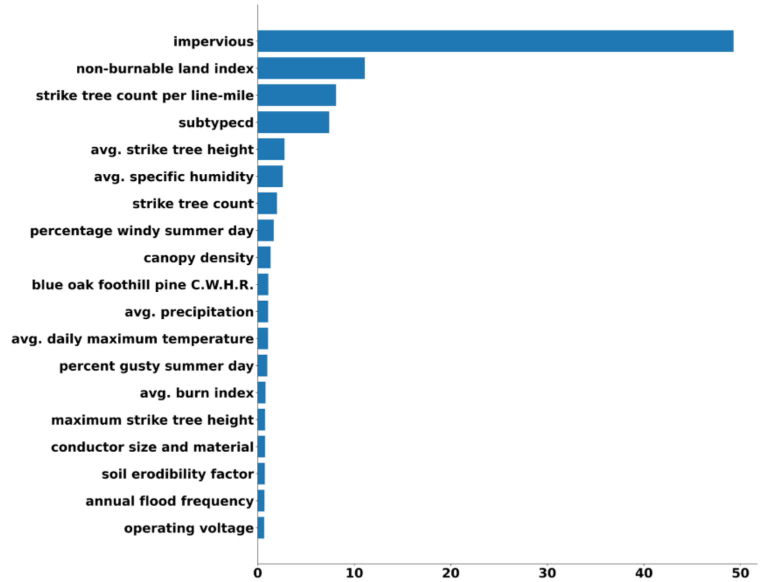

Figure 63 – Third-Party – Balloon Top Feature Importance

There are numerous additional covariate features that marginally contribute to the Balloon model, accounting for the remaining 5% of model performance:

- maximum wind speed
- tier-2
- avg. vapor pressure deficit
- P.S.P.S. segment
- soil horizon thickness
- soil clay content
- closed cone pine cypress CWHR
- sierran mixed conifer CWHR
- red fir CWHR
- soil slope
- soil organic matter
- avg. energy release
- redwood CWHR
- tier-3
- soil drain
- soil hydro-group
- available soil water capacity
- is soil hydric
- minimum strike tree height
- eucalyptus C.W.H.R
- avg. wind speed
- tree species
- soil liquid limit
- valley oak woodland CWHR
- montane riparian CWHR

3.5.2.2.1.6 Limitations and Opportunities

Due to population dynamics, it is important that the Third-Party models be able to capture temporal variations over the years and project into the future based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends nor can it see finer trends such as demographic change and (de)urbanization. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.2.2 Third-Party – Vehicle Model

The Third-Party Vehicle model uses historical vehicle-caused failure events to model outage probability from vehicle contact with electrical assets resulting in damage to conductors.

3.5.2.2.2.1 Key Developments

The only significant change to the Vehicle model for v4 was that the failure event data set was updated to include 2023 failures, increasing the number of events used to for modeling training.

3.5.2.2.2.2 Performance

The v4 Third-Party – Vehicle outage model demonstrated very good model performance with an AUC of 0.76, as shown in *Figure 64*.

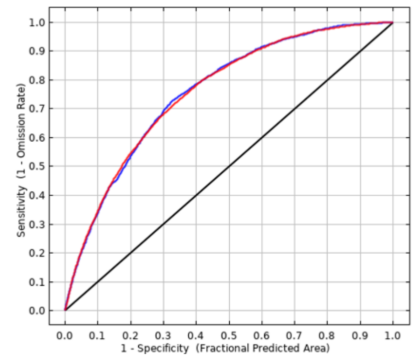

Figure 64 - Third Party-Vehicle ROC Curve

3.5.2.2.2.3 Model Data

Table 24 presents a summary of event counts used for v4 development:

Table 24 – Third-Party – Vehicle Event Summary

Event Type	Value
Outage Count	8347
Ignition Count	305
Ignitions Per Outage	0.0365

3.5.2.2.2.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.2.2.5 Covariate Selection

Figure 65 presents the total gains feature importance that account for 95% of the Third-Party - Vehicle model performance. Top features include the impervious and non-burnable land indices that indicate urban environments, strike tree count, conductor type (subtypecd), and conductor size and material.

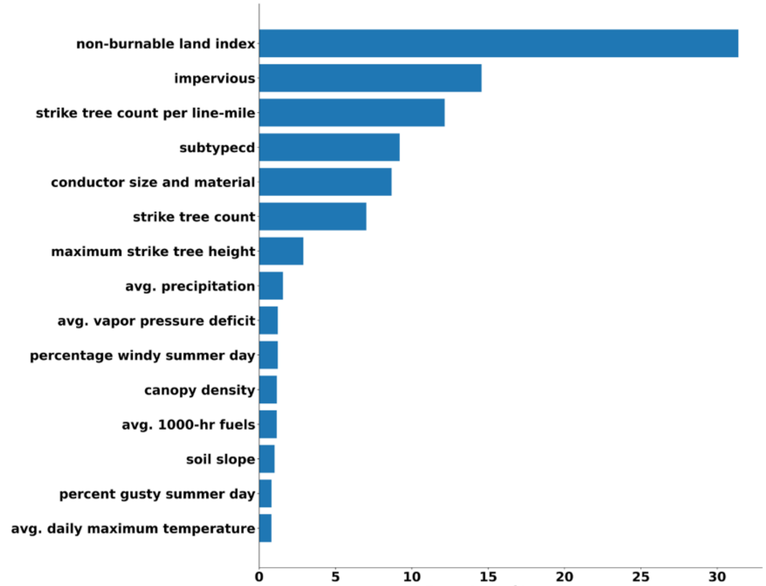

Figure 65 - Third Party-Vehicle Top Feature Performance

There are numerous additional covariate features that marginally contribute to the Vehicle model, accounting for the remaining 5% of model performance:

- soil permeability
- avg. strike tree height
- local topography
- avg. burn index
- operating voltage
- tree species
- soil hydro-group
- P.S.P.S. segment
- ponderosa pine CWHR
- coastal oak woodland CWHR
- soil liquid limit
- available soil water capacity
- montane hardwood CWHR
- tier-3
- is soil hydric
- blue oak foothill pine CWHR
- soil drain
- juniper CWHR
- avg. wind speed
- blue oak woodland CWHR
- soil erodibility factor
- sierran mixed conifer CWHR
- eucalyptus C.W.H.R
- avg. 100-hr fuels
- montane hardwood CWHR
- closed cone pine cypress CWHR
- annual soil flood frequency
- tier-2
- soil horizon thickness
- valley oak woodland CWHR
- soil clay content
- redwood CWHR

3.5.2.2.2.6 Limitations and Opportunities

Due to population dynamics, it is important that the Third-Party models be able to capture temporal variations over the years and project into the future based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends nor can it see finer trends such as demographic change and (de)urbanization. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.2.3 Third-Party – Other Model

The Third-Party – Other model uses historical third-party failure events that are not classified as mylar balloon or vehicle-related failures.

3.5.2.2.3.1 Key Developments

The only significant change to the Third-Party – Other model for v4 was that the failure event data set was updated to include 2023 failures, increasing the number of events used to for modeling training.

3.5.2.2.3.2 Performance

The v4 Third-Party – Other outage model demonstrated good model performance with an AUC of 0.74, as shown in *Figure 66*.

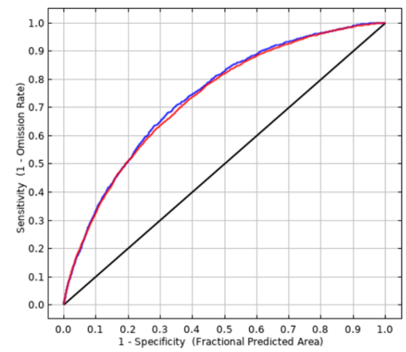

Figure 66 - Third Party-Other ROC Curve

3.5.2.2.3.3 Model Data

Table 25 presents a summary of event counts used for v4 development:

Table 25 - Third-Party – Other Event Summary

Event Type	Value
Outage Count	2785
Ignition Count	151
Ignitions Per Outage	0.0542

3.5.2.2.3.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4 due to its ability to analyze presence only events and preserve spatial interactions during model optimization.

3.5.2.2.3.5 Covariate Selection

Figure 67 presents the total gains feature importance that account for 95% of the Third-Party - Other model performance. Top features include the impervious and non-burnable land indices that indicate urban environments, strike tree count, conductor type (subtypecd), and strike tree height.

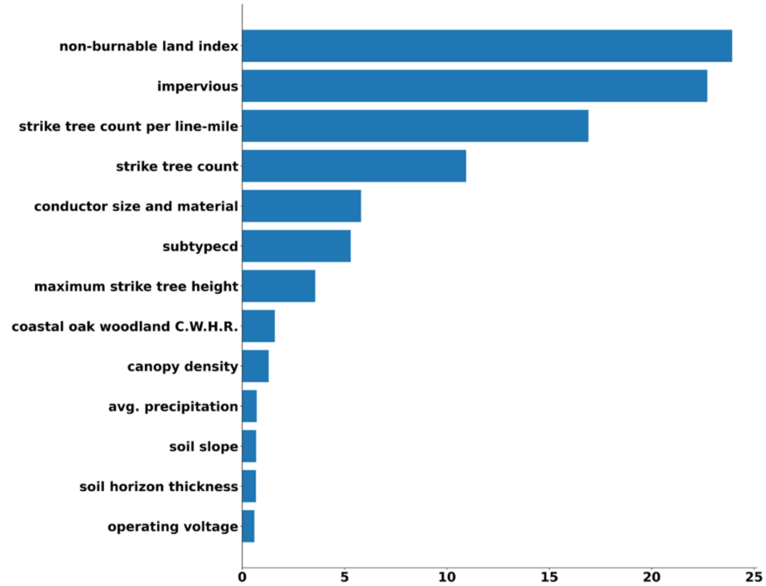

Figure 67 - Third Party-Other Top Feature Performance

There are numerous additional covariate features that marginally contribute to the Other model, accounting for the remaining 5% of model performance:

- avg. specific humidity
- avg. strike tree height
- windy summer day percentage
- avg. vapor pressure deficit
- annual soil flood frequency
- tier-3
- tier-2
- avg. daily maximum temperature
- soil liquid limit
- Klamath mixed conifer CWHR
- local topography
- valley foothill riparian CWHR
- is soil hydric
- pinyon juniper CWHR
- blue oak woodland CWHR
- soil drain
- soil permeability
- blue oak foothill pine CWHR
- P.S.P.S. segment
- soil hydro-group
- gusty summer day percentage
- montane hardwood CWHR
- soil clay content
- sierran mixed conifer CWHR
- available soil water capacity
- juniper CWHR
- avg. wind speed
- eucalyptus C.W.H.R
- desert riparian CWHR
- ponderosa pine CWHR
- avg. energy release
- maximum wind speed

3.5.2.2.3.6 Limitations and Opportunities

Due to population dynamics, it is important that the Third-Party models be able to capture temporal variations over the years and project into the future based on current trends. The current MaxEnt algorithm, while very good at modeling spatial trends, cannot model temporal trends nor can it see finer trends such as demographic change and (de)urbanization. Deep learning models such as Recurrent Neural Networks and Transformers will be investigated to test if adding the temporal domain to future models will improve model predictive performance.

3.5.2.3 Vegetation Models

Vegetation regularly comes into contact with electrical assets and often initiates failure, outages, and ignitions. Historically, trees, through two causal paths, are responsible for majority of vegetation-related failures and outages. Therefore, vegetation outages are predicted using three distinct models:

- Branch – Failures caused by tree branches striking conductors.
- Trunk – Failures caused by full tree body striking conductors.
- Other – Failures from other or unknown vegetation causes.

3.5.2.3.1 Vegetation – Branch Model

The Vegetation – Branch model uses historical branch failure events to model outage probability from branch debris striking a conductor.

3.5.2.3.1.1 Key Developments

During v4 development, we explored the issue of bark beetle infestation as a source of tree failures. A study was conducted to discern the cause of the infestations and it was found that the lack of water availability over the years, primarily due to Climate Change, was hampering the tree's immune system and natural defenses. These effects were captured by way of the following remotely sensed quantities and the surveyed tree mortality map by CalFIRE:

- The Standardized Precipitation and Evapo-Transpiration Index – An index used to quantify regional water availability.
- The Climatic Water Deficit – This quantity represents the water use of trees in the area.
- California High Hazard Zones – This CalFIRE product was used to give the model a sense of the tree mortality rate in any given area.

Other important model developments for v4 include:

- Planet Labs Tree Canopy Density – Similar to counting the number of strike trees around an asset, a measure of the canopy density was used to quantify the amount of material that was available to be ejected from any given tree.
- Wind Strike Potential – New attributes were introduced to consider the relative position of trees to conductors along with wind speed and direction to when predicting branch strike potential.
- Tree Failure Locations – Tree failures were attributed to actual tree locations for v4, whereas v3 used the locations of triggered protective devices.

3.5.2.3.1.2 Performance

The v3 Branch model lagged the Trunk model in performance. The addition of the new covariates improved the v4 Branch model and it now matches the Trunk Model in performance. The v4 Vegetation – Branch outage model demonstrated excellent model performance with an AUC of 0.87, as shown in *Figure 68*.

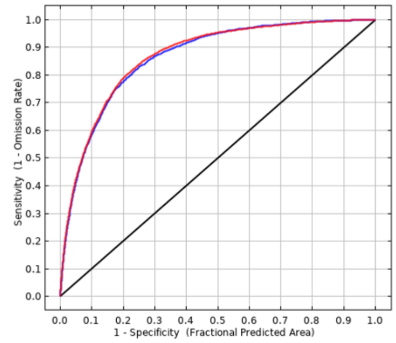

Figure 68 - Vegetation-Branch ROC Curve

3.5.2.3.1.3 Model Data

Table 26 presents a summary of event counts used for v4 development:

Table 26 – Vegetation - Branch Event Summary

Event Type	Value
Outage Count	7392
Ignition Count	428
Ignitions Per Outage	0.058

In addition to updated target events, newly added covariates were also generated by various base data:

- The Standardized Precipitation and Evapo-transpiration Index (SPEI) represents water availability. This index measures the deviation of drought conditions from a multi-year baseline quantifying the onset, duration, and magnitude of drought.
- The Climatic Water Deficit (CWD) measures vegetation water use. The CWD data comes from the USGS Water Basin Characterization Model.
- As a proxy for tree mortality, both tier-1 and tier-2 CalFIRE California High Hazard Zones were implemented as covariates.
- Canopy Density was implemented from a dataset from Planet Labs.
- The following new wind covariates were developed:
  - The normalized historical wind direction and strike potential
  - The historical wind direction and strike potential
  - The wind damage and strike potential
  - The normalized wind damage and strike potential which measure different ways in which a wind’s magnitude and direction interplay with a tree’s location and a conductor’s location to ultimately propel matter into a conductor.

3.5.2.3.1.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4. A significant issue for the vegetation models is inability to determine a target probability of a vegetation failure causing damage to an electrical asset. Unfortunately, a vegetation failure is only recorded as failure if it impacts an electrical asset. Therefore, no information is available about how many benign vegetation failures occur that do not result in an asset problem. The MaxEnt algorithm is particularly suited for modeling vegetation failures as it only requires knowledge of known asset failure locations caused by falling branches or tree trunks and does not require data on failures that did not result in an asset impact.

3.5.2.3.1.5 Covariate Selection

Figure 69 presents the total gains feature importance that account for 95% of the Vegetation - Branch model performance. Top features include strike tree counts, strike tree height, the impervious and non-burnable land indices that indicate urban environments, and conductor (subtypecd).

There are numerous additional covariate features that marginally contribute to the Branch model, accounting for the remaining 5% of model performance:

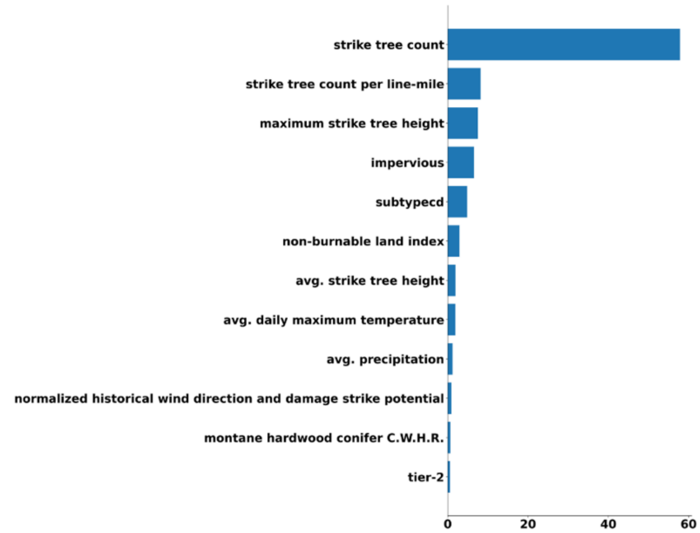

Figure 69 - Vegetation-Branch Top Feature Performance

- blue oak foothill pine CWHR
- operating voltage
- wind damage and strike potential
- coastal
- soil hydro-group
- maximum 4-yr CWD
- valley foothill riparian CWHR
- douglas fir CWHR
- soil permeability
- avg. specific humidity
- soil slope
- gusty summer day percentage
- historical wind direction strike pot ential
- annual soil flood frequency
- soil drain
- coastal oak woodland CWHR
- canopy density
- avg. 20-yr S.P.E.I.
- normalized historical wind directi on strike potential
- local topography
- is soil hydric
- windy summer day percentage
- closed cone pine cypress CWHR
- coefficient of variation 4-yr CWD
- avg. energy release
- avg. wind speed
- avg. 100-hr fuels
- tree species
- avg. vapor pressure deficit
- soil liquid limit
- tier-2 tree mortality H.H.Z.
- soil horizon thickness
- soil clay content
- lodgepole pine CWHR
- soil erodibility factor
- blue oak woodland CWHR
- eucalyptus C.W.H.R
- maximum 20-yr CWD

3.5.2.3.1.6 Limitations and Opportunities

There are two significant opportunities to strengthen the vegetation models for future releases.

The first opportunity is to extend our modeling into the temporal domain. As vegetation grows and changes over the years, it is important to capture this change along with its interplay with a changing climate. Deep learning architectures such as Recurrent Neural Networks and Transformers will be explored in the future to include this temporal dimension. In addition, the use of Planet Labs data products will be used to model tree conditions from 2016 onward to accurately reflect the changing vegetation conditions from growth, failure, and mitigation efforts by vegetation management.

The second area where improvement can be made is that tree mortality is not accurately captured by the CalFIRE dataset. Yearly tree mortality data has been purchased from Planet Labs, and is being prepared for use in future models. It is expected that the new tree mortality data will be included for the next version of the vegetation models.

3.5.2.3.2 Vegetation – Trunk Model

The Vegetation – Trunk model uses historical tree body, or trunk, failure events to model outage probability from tree falls striking a conductor.

3.5.2.3.2.1 Key Developments

During v4 development, we explored the issue of bark beetle infestation as a source of tree failures. A study was conducted to discern the cause of the infestations and it was found that the lack of water availability over the years, primarily due to Climate Change, was hampering the tree's immune system and natural defenses. These effects were captured by way of the following remotely sensed quantities and the surveyed tree mortality map by CalFIRE:

- The Standardized Precipitation and Evapo-Transpiration Index – An index used to quantify regional water availability.
- The Climatic Water Deficit – This quantity represents the water use of trees in the area.
- California High Hazard Zones – This CalFIRE product was used to give the model a sense of the tree mortality rate in any given area.

Other important model developments for v4 include:

- Planet Labs Tree Canopy Density – Similar to counting the number of strike trees around an asset, a measure of the canopy density was used to quantify the amount of material that was available to be ejected from any given tree.
- Wind Strike Potential – New attributes were introduced to consider the relative position of trees to conductors along with wind speed and direction to when predicting branch strike potential.
- Tree Failure Locations – Tree failures were attributed to actual tree locations for v4, whereas v3 used the locations of triggered protective devices.

3.5.2.3.2.2 Performance

The v4 Vegetation – Trunk outage model demonstrated excellent model performance with an AUC of 0.88, as shown in *Figure 70*.

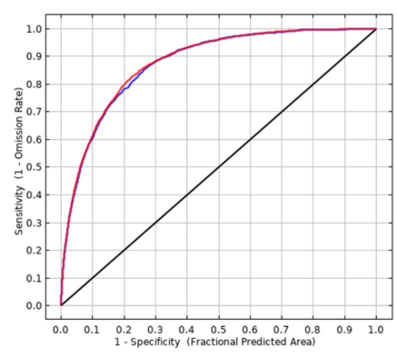

Figure 70 - Vegetation-Trunk ROC Curve

3.5.2.3.2.3 Model Data

Table 27 presents a summary of event counts used for v4 development:

Table 27 - Vegetation - Trunk Event Summary

Event Type	Value
Outage Count	4630
Ignition Count	344
Ignitions Per Outage	0.074

In addition to updated target events, newly added covariates were also generated by various base data:

- The Standardized Precipitation and Evapo-transpiration Index (SPEI) represents water availability. This index measures the deviation of drought conditions from a multi-year baseline quantifying the onset, duration, and magnitude of drought.
- The Climatic Water Deficit (CWD) measures vegetation water use. The CWD data comes from the USGS Water Basin Characterization Model.
- As a proxy for tree mortality, both tier-1 and tier-2 CalFIRE California High Hazard Zones were implemented as covariates.
- Canopy Density was implemented from a dataset from Planet Labs.
- The following new wind covariates were developed:
  - The normalized historical wind direction and strike potential
  - The historical wind direction and strike potential
  - The wind damage and strike potential
  - The normalized wind damage and strike potential which measure different ways in which a wind's magnitude and direction interplay with a tree's location and a conductor's location to ultimately propel matter into a conductor.

3.5.2.3.2.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4. A significant issue for the vegetation models is inability to determine a target probability of a vegetation failure causing damage to an electrical asset. Unfortunately, a vegetation failure is only recorded as failure if it impacts an electrical asset. Therefore, no information is available about how many benign vegetation failures occur that do not result in an asset problem. The MaxEnt algorithm is particularly suited for modeling vegetation failures as it only requires knowledge of known asset failure locations caused by falling branches or tree trunks and does not require data on failures that did not result in an asset impact.

3.5.2.3.2.5 Covariate Selection

Figure 71 presents the total gains feature importance that account for 95% of the Vegetation - Trunk model performance. Top features include strike tree counts, strike tree height, and subtypecd.

There are numerous additional covariate features that marginally contribute to the Trunk model, accounting for the remaining 5% of model performance:

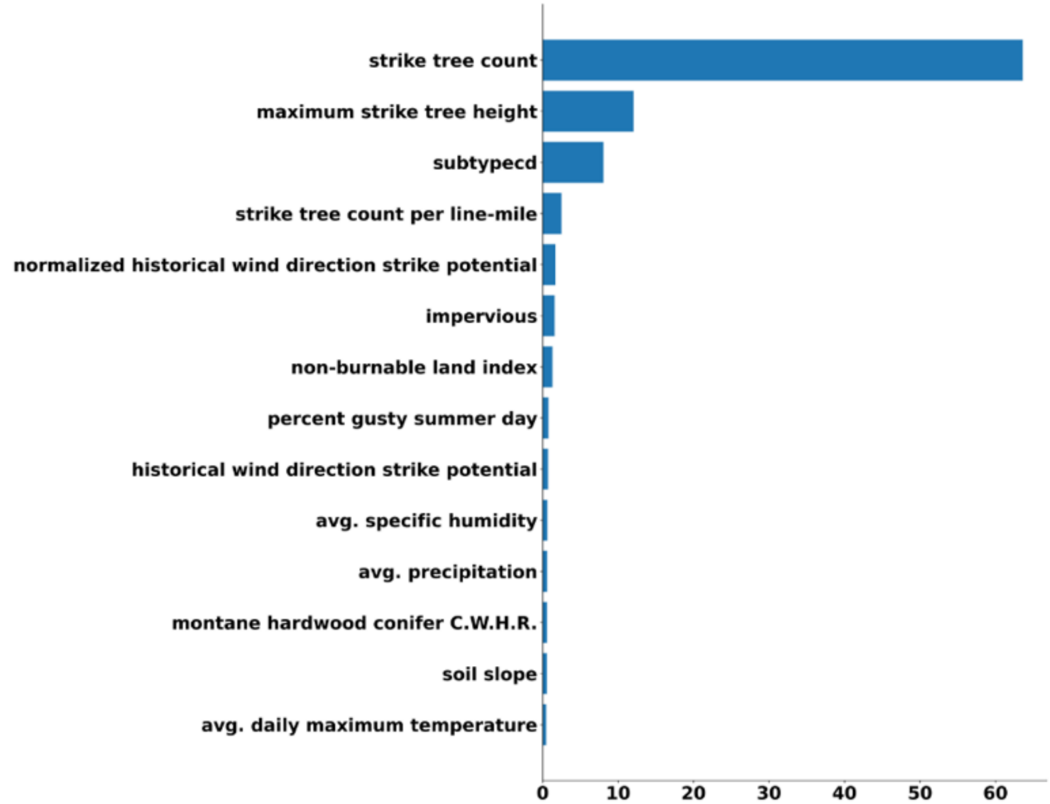

Figure 71 - Vegetation-Trunk Top Feature Performance

- avg. 1000-hr fuels
- avg. vapor pressure deficit
- soil hydro-group
- eucalyptus CWHR
- maximum 20-yr CWD
- canopy density
- soil clay content
- avg. wind speed
- normalized wind damage and str ike potential
- soil permeability
- tier-2 tree mortality H.H.Z.
- blue oak foothill pine CWHR
- juniper CWHR
- wind damage / strike potential
- avg. 20-yr S.P.E.I.
- coefficient of variation 4-yr CWD
- coefficient of variation 20-yr CW D
- tree species
- soil horizon thickness
- avg. burn index
- local topography
- white fir CWHR
- operating voltage
- closed cone pine cypress CWHR
- valley foothill riparian CWHR
- windy summer day percentage
- coastal oak woodland CWHR
- montane riparian CWHR
- avg. strike tree height
- avg. energy release
- soil erodibility factor
- sierran mixed conifer CWHR
- red fir CWHR
- soil drain
- maximum 4-yr CWD
- pinyon juniper CWHR
- is soil hydric
- tier-3
- ponderosa pine CWHR
- sub-alpine conifer CWHR
- coastal
- maximum 20-yr S.P.E.I.
- douglas fir CWHR

3.5.2.3.2.6 Limitations and Opportunities

There are two significant opportunities to strengthen the vegetation models for future releases.

The first opportunity is to extend our modeling into the temporal domain. As vegetation grows and changes over the years, it is important to capture this change along with its interplay with a changing climate. Deep learning architectures such as Recurrent Neural Networks and Transformers will be explored in the future to include this temporal dimension. In addition, the use of Planet Labs data products will be used to model tree conditions from 2016 onward to accurately reflect the changing vegetation conditions from growth, failure, and mitigation efforts by vegetation management.

The second area where improvement can be made is that tree mortality is not accurately captured by the CalFIRE dataset. Yearly tree mortality data has been purchased from Planet Labs, and is being prepared for use in future models. It is expected that the new tree mortality data will be included for the next version of the vegetation models.

3.5.2.3.3 Vegetation – Other Model

The Vegetation – Other model uses historical failure events related to third-party, snow load, wind, and unclassified events to model outage probability.

3.5.2.3.3.1 Key Developments

During v4 development, we explored the issue of bark beetle infestation as a source of tree failures. A study was conducted to discern the cause of the infestations and it was found that the lack of water availability over the years, primarily due to Climate Change, was hampering the tree's immune system and natural defenses. These effects were captured by way of the following remotely sensed quantities and the surveyed tree mortality map by CalFIRE:

- The Standardized Precipitation and Evapo-Transpiration Index – An index used to quantify regional water availability.
- The Climatic Water Deficit – This quantity represents the water use of trees in the area.
- California High Hazard Zones – This CalFIRE product was used to give the model a sense of the tree mortality rate in any given area.

Other important model developments for v4 include:

- Planet Labs Tree Canopy Density – Similar to counting the number of strike trees around an asset, a measure of the canopy density was used to quantify the amount of material that was available to be ejected from any given tree.
- Wind Strike Potential – New attributes were introduced to consider the relative position of trees to conductors along with wind speed and direction to when predicting branch strike potential.
- Tree Failure Locations – Tree failures were attributed to actual tree locations for v4, whereas v3 used the locations of triggered protective devices.

3.5.2.3.3.2 Performance

The v4 Vegetation – Other outage model demonstrated excellent model performance with an AUC of 0.88, as shown in *Figure 72*.

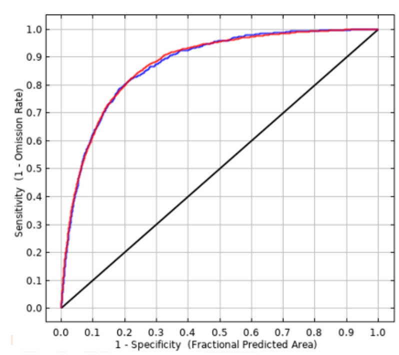

Figure 72 - Vegetation-Other ROC Curve

3.5.2.3.3.3 Model Data

Table 28 presents a summary of event counts used for v4 development:

Table 28 - Vegetation - Other Event Summary

Event Type	Value
Outage Count	1931
Ignition Count	233
Ignitions Per Outage	0.115

3.5.2.3.3.4 Algorithm Selection

The MaxEnt algorithm, which was used for v3, was retained for v4. A significant issue for the vegetation models is inability to determine a target probability of a vegetation failure causing damage to an electrical asset. Unfortunately, a vegetation failure is only recorded as failure if it impacts an electrical asset. Therefore, no information is available about how many benign vegetation failures occur that do not result in an asset problem. The MaxEnt algorithm is particularly suited for modeling vegetation failures as it only requires knowledge of known asset failure locations caused by falling branches or tree trunks and does not require data on failures that did not result in an asset impact.

3.5.2.3.3.5 Covariate Selection

Figure 73 presents the total gains feature importance that account for 95% of the Vegetation - Other model performance. Top features include strike tree counts, the impervious and non-burnable land indices that indicate urban environments, subtypecd, and strike tree height.

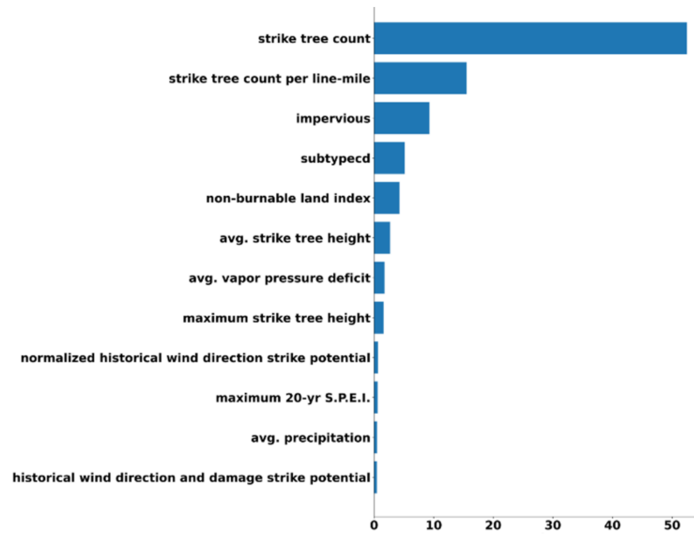

Figure 73 - Vegetation-Other Top Feature Performance

There are numerous additional covariate features that marginally contribute to the Other model, accounting for the remaining 5% of model performance:

- tier-2 tree mortality H.H.Z.
- avg. burn index
- operating voltage
- Coefficient of variation 20-yr S.P. E.I.
- coastal oak woodland CWHR
- blue oak woodland CWHR
- available soil water capacity
- soil hydro-group
- avg. 20-yr S.P.E.I.
- soil erodibility factor
- tier-3
- avg. energy release
- maximum wind speed
- soil permeability
- canopy density
- closed cone pine cypress CWHR
- historical wind direction strike pote ntial
- windy summer day percentage
- tree species
- soil slope
- soil organic matter
- ponderosa pine CWHR
- avg. specific humidity
- annual soil flood frequency
- avg. 1000-hr fuels
- soil liquid limit
- gusty summer day percentage
- valley foothill riparian CWHR
- klamath mixed conifer CWHR
- local topography
- montane riparian CWHR
- normalized wind damage and strike potential
- soil clay content
- juniper CWHR
- jeffrey pine CWHR
- montane hardwood CWHR
- minimum 20-yr S.P.E.I.
- redwood CWHR
- avg. wind speed

3.5.2.3.3.6 Limitations and Opportunities

There are two significant opportunities to strengthen the vegetation models for future releases.

The first opportunity is to extend our modeling into the temporal domain. As vegetation grows and changes over the years, it is important to capture this change along with its interplay with a changing climate. Deep learning architectures such as Recurrent Neural Networks and Transformers will be explored in the future to include this temporal dimension. In addition, the use of Planet Labs data products will be used to model tree conditions from 2016 onward to accurately reflect the changing vegetation conditions from growth, failure, and mitigation efforts by vegetation management.

The second area where improvement can be made is that tree mortality is not accurately captured by the CalFIRE dataset. Yearly tree mortality data has been purchased from Planet Labs and is being prepared for use in future models. It is expected that the new tree mortality data will be included for the next version of the vegetation models.

3.6 Probability of Ignition Model

The probability of ignition p(i) is the likelihood that an asset-based ignition will occur during a fire season. Probability of ignition is predicted by the Probability of Ignition Given Outage model using the probability of outage predictions from all of the Asset Equipment and Contact From Object models along with other attributes such as environmental conditions. Fire season probability of ignitions are individually predicted for each specific Asset Equipment and Contact From Object model.

3.6.1 Probability of Ignition Given Outage

The Probability of Ignition Given Outage model, p(i|o), takes as its input the results from an Asset Equipment or Contact From Object model, p(o). The percentage of outages that result in an ignition varies, as one would expect, on the outage type. The p(i|o) model uses failure model-specific attributes and environmental conditions to determine the likelihood that a given outage is likely to result in an ignition.

For asset-based event models, the probability of ignition for a given asset is the product of its probability of ignition given an outage and its probability of an outage:

$$p(i)_{asset} = p(i|o)_{asset} * p(o)_{asset}$$

For Contact From Object models, which are location, pixel-based models, the probability of ignition for a given location is the product of the location probability of ignition given outage and the location probability of outage for a specific model:

$$p(i)_{loc} = p(i|o)_{loc} * p(o)_{loc}$$

Individual asset and contact from object probabilities can be composited to determine a summed probability of ignition for a location:

$$p(i)_{loc} = \sum_{loc/asset} p(i|o)_{loc/asset} * p(o)_{loc/asset}$$

Where:

asset	Modeled asset type: conductor, transformer, support structure, etc.
loc	Asset location expressed as a 100m square pixel within the PG&E service territory
p(i)	Probability of Ignition
p(i|o)	Probability of Ignition given an Outage
p(o)	Probability of Outage

3.6.1.1 Key Developments

The p(i|o) model developed for the DEPM v4 provides two substantial improvements over the v3 model:

- Event cause and equipment type interaction terms
- EPSS compensation

The p(i|o) model correlates to failure/outage rates, weather conditions, fuel conditions and availability, and other location-specific attributes. However, the correlation between fuel and weather conditions and ignition outcomes also depends on the nature of the underlying events. Specifically, some events, like transformer failures predominantly result in pole fires that are not influenced by fuels on the ground, while others, like insulator tracking faults, require moisture and condensation for an event to occur. The introduction of event cause and equipment type labels for events allowed the use of interaction terms that produce separate weather and fuels correlation terms for distinct groups of events that share the same characteristics.

An important purpose for the p(i|o) model is to support tradeoffs between mitigation strategies, including EPSS. Therefore, the v4 p(i|o) model needed to be calibrated to predict the number of ignitions that would be expected without EPSS. Section *3.3* explains how the ignitions event training data was modified to account for EPSS impacts.

3.6.1.2 Performance

Table 29 provide an overview of the p(i|o) performance for the entire suite of probability of failure/outage models. The important outcome to understand from this table is whether the p(i|o) model improves the ability to predict ignitions against the assumption that you can directly predict an ignition from the prediction of a failure/outage. The Delta AUC column highlights the p(i|o) improvement for predicting ignitions, where a positive number is an improvement, a negative number is a degradation. Note that the two models that show the most degradation using the p(i|o) model are also low performing probability of failure/outage models.

Table 29 – p(i|o) Predictive Performance Gain

Event Model	p(o) AUC	p(i) AUC	Delta AUC	p(o) Conc. Factor	p(i) Conc. Factor
Animal – Bird	0.651	0.683	**0.032**	1.79	2.16
Animal – Squirrel	0.839	0.852	0.013	3.70	3.91
Animal - Other	0.741	0.756	0.015	2.53	2.93
Capacitor Bank	0.558	0.549	-0.009	1.40	1.40
DPD	0.625	0.475	-0.151	1.00	1.50
Fuse	0.669	0.681	0.013	2.00	2.10
Primary Conductor – Line Slap	0.770	0.757	-0.013	3.33	3.33
Primary Conductor – Wire Down	0.628	0.639	0.010	1.91	1.78
Primary Conductor - Other	0.694	0.706	0.013	2.75	2.50
Secondary Conductor	0.718	0.728	0.010	2.49	2.59
Support Structure - Electrical	0.854	0.848	-0.006	3.90	3.70
Support Structure - Equipment	0.602	0.654	**0.052**	1.63	2.02
Switch	0.637	0.608	-0.029	1.67	1.67
Third-Party - Balloon	0.801	0.817	0.016	3.25	3.42
Third-Party - Vehicle	0.681	0.712	**0.031**	1.85	2.23
Third-Party - Other	0.613	0.672	**0.060**	1.78	2.05
Transformer – Equipment	0.725	0.748	0.023	2.94	3.53
Transformer – Leaking	0.865	0.845	-0.019	5.00	5.00
Vegetation - Branch	0.804	0.824	0.020	3.20	3.43
Vegetation - Trunk	0.825	0.832	0.007	3.40	3.45
Vegetation - Other	0.760	0.771	0.010	2.57	2.74
Voltage Regulator	0.739	0.720	-0.019	2.00	2.00
Other Equipment	0.640	0.650	0.010	1.82	2.05

The p(i|o) model was fit using a random selection of 75% of the target set data, with the remaining 25% set aside for predictive performance testing. The model classification performance is illustrated in *Figure 74* showing the training and testing ROC curves. The ROC curve indicates the degree to which model assigns elevated ignition probabilities to outages that were associated with real ignitions. The curve demonstrates that the p(i|o) model performs very well, capturing 80% of true ignitions after covering just under 25% of the target set in rank order.

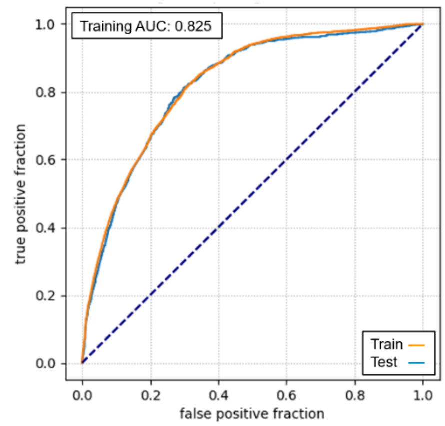

Figure 74 - p(i|o) ROC Curve

3.6.1.3 Model Data

The various outage types predicted by the Asset Equipment and Contact From Object event models have different physical pathways to ignition and, therefore, different rates of ignitions as shown in *Table 30*. A few examples of different ignition routes include:

- Conductor and tree trunk failures often lead to wire down events, placing fault current directly in contact with ground fuels.
- Conductor line slap and branch contacts may cause phase faults, with arcing or combustion that can lead to molten metal, sparks, or flaming branches falling onto ground fuels.
- Equipment tracking, with current arcing across damp, dirty equipment, can often lead to pole fires, which in many cases fail to ignite ground fuels at all.

Table 30 – Outage and Ignition Event Counts Summary

Event Model	Outages (count)	Ignitions (count)	Ignitions per Outage
Animal – Bird	5,429	259	4.77%
Animal – Squirrel	4,488	51	1.14%
Animal - Other	1,291	105	8.13%
Capacitor Bank	370	125	**33.78%**
DPD	624	70	11.22%
Fuse	4,474	245	5.48%
Primary Conductor – Line Slap	938	39	4.16%
Primary Conductor – Wire Down	5,317	553	10.40%
Primary Conductor - Other	1,961	191	9.74%
Secondary Conductor	3,318	223	6.72%
Support Structure - Electrical	4,033	**835**	20.70%
Support Structure - Equipment	4,314	212	4.91%
Switch	790	59	7.47%
Third-Party - Balloon	2,420	120	4.96%
Third-Party - Vehicle	8,347	305	3.65%
Third-Party - Other	2,785	151	5.42%
Transformer – Equipment	14,297	292	2.04%
Transformer – Leaking	753	**1**	**0.13%**
Vegetation - Branch	7,392	428	5.79%
Vegetation - Trunk	4,630	344	7.43%
Vegetation - Other	1,931	223	11.55%
Voltage Regulator	**286**	29	10.14%
Other Equipment	**54,053**	201	0.37%
All Event Models	**134,241**	**5,061**	**3.77%**

3.6.1.4 Algorithm Selection

The v4 p(i|o) model, like the v3 version, is a Logistic Regression model. Random Forest models were also examined but were found to produce similar predictive power but with worse interpretability. Logistic Regression is well-suited for isolating the impact of weather and fuel conditions on ignition rates and supports a semi-causal interpretation of model results.

The introduction of event cause and equipment type interaction terms for v4 required the implementation of a more sophisticated feature generation and selection approach, not just for model training, but also for predicting at all locations across all representative days to produce marginalized expected annual ignition

3.6.1.5 Covariate Selection

The multiple paths to ignition events are influenced differently by environmental conditions such as fuel moisture and wind. Therefore, the p(i|o) model must reflect that the outages predicted by each outage event model have their own unique interaction with environmental variables.

Unfortunately, from a modeling perspective, many of the event models do not have enough historical ignitions that can be used for p(i|o) model training. The lack of sufficient training data makes it likely that a model would be over-fit and subsequently perform poorly against test data. To compensate, outage event models with similar ignition behaviors in response to environmental conditions were grouped together to improve p(i|o) model performance. Optimal performance was achieved using the model groups defined in *Table 31*.

Table 31 - p(i|o) Modeling Groups

Model Group	Member Models	Outages (count)	Ignitions (count)	Ignitions per Outage
1	Vegetation – Branch, Vegetation – Trunk	12,022	751	6.2%
2	Support Structure – Electrical	4,033	713	17.7%
3	Animal – Squirrel, Third-Party – Balloon, Third-Party – Vehicle, Third-Party – Other	18,040	613	3.4%
4	Primary Conductor – Wire Down	5,317	515	9.7%
5	DPD, Capacitor Bank, Fuse, Switch, Voltage Regulator	6,544	498	7.6%
6	Animal – Bird, Animal – Other, Primary Conductor – Line Slap, Transformer – Leaking	8,411	388	4.6%
7	Transformer – Equipment	14,297	269	1.9%
8	Secondary Conductor	3,318	220	6.6%
9	Vegetation – Other	1,931	209	10.8%
10	Other Equipment	54,053	188	0.3%
11	Support Structure – Equipment	4,314	187	4.3%
12	Primary Conductor – Other	1,961	179	9.1%

Many combinations of model-grouped and ungrouped attributes were tested as features in the p(i|o) model. *Figure 75* presents the total gains feature importance for the highest performing features in the p(i|o) model.

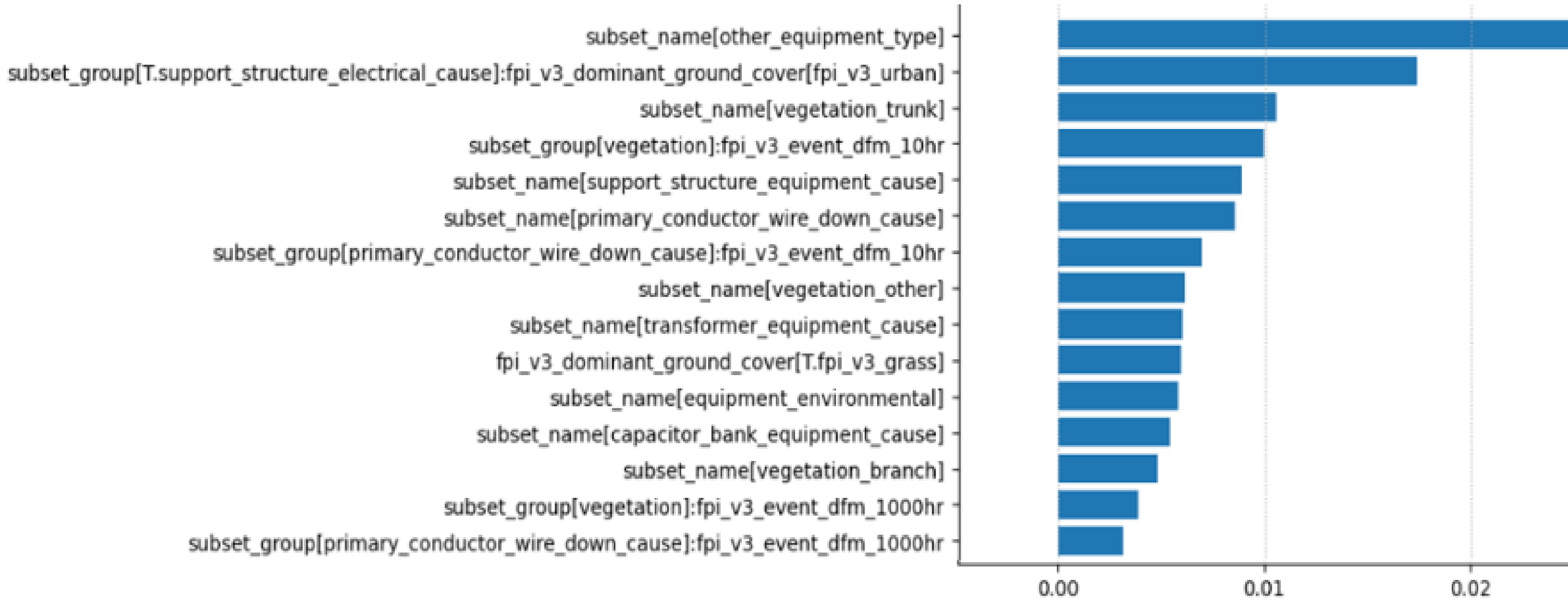

Figure 75 - p(i|o) Top Feature Importance

Other covariate features that moderately contribute to the p(i|o) model include:

- subset_group [vegetation other]: fpi v3 event dfm 10hr
- subset_group [T.support structure equipment]: fpi v3 dominant ground cover [fpi v3 urban]
- subset_group [T.vegetation]: fpi v3 dominant ground cover [fpi v3 grass shrub]
- subset_name [animal squirrel]
- subset_group [other2]: fpi v3 event dfm 1000hr
- fpi_dominant_ground_cover[T.fpi v3 shrub]
- subset_group [support structure electrical]: fpi v3 event dfm 1000hr
- subset_group [other2]: fpi v3 event dfm 10hr
- subset_group [T.other2]: fpi v3 dominant ground cover [fpi v3 grass shrub
- subset_name [transformer leaking]
- subset_group [T.vegetation other]: fpi v3 dominant ground cover [fpi v3 grass shrub]
- subset_group [T.other equipment]: fpi v3 dominant ground cover [fpi v3 grass]
- subset_group [T.protection actuation]: fpi v3 dominant ground cover [fpi v3 grass]
- subset_group [support structure equipment]: fpi v3 event dfm 10hr

- subset_group [T.support structure electrical]: fpi v3 dominant ground cover [fpi v3 timber litter]
- subset_group [T.transformer equipment]: fpi v3 dominant ground cover [fpi v3 timber litter]
- subset_group [secondary conductor]: fpi v3 event dfm 10hr
- subset_group [other equipment]: fpi v3 event dfm 10hr
- subset_group [primary conductor other]: fpi v3 event dfm 1000hr
- subset_group [sother1]: fpi v3 event dfm 10hr
- subset_group [T.transformer equipment]: fpi v3 dominant ground cover [fpi v3 grass]]
- subset_group [T.primary conductor other]: fpi v3 dominant ground cover [fpi v3 grass shrub]
- subset_name [DPD]
- subset_group [T.primary conductor wire down]: fpi v3 dominant ground cover [fpi v3 grass shrub]
- subset_group [other1]: fpi v3 event dfm 1000hr
- subset_group [T.secondary conductor]: fpi v3 dominant ground cover [fpi v3 grass]
- subset_group [vegetation other]: fpi v3 event dfm 1000hr
- subset_group [T.support structure equipment]: fpi v3 dominant ground cover [fpi v3 shrub]
- subset_group [T.transformer equipment]: fpi v3 dominant ground cover [fpi v3 shrub]
- subset_group [transformer equipment]: fpi v3 event dfm 1000hr
- subset_group [transformer equipment]: fpi v3 event dfm 10hr
- subset_group [T.primary conductor other]: fpi v3 dominant ground cover [fpi v3 timber understory]
- subset_group [T.support structure equipment]: fpi v3 dominant ground cover [fpi v3 timber understory]
- subset_group [T.vegetation other]: fpi v3 dominant ground cover [fpi v3 shrub]
- subset_group [support structure electrical]: fpi v3 event dfm 10hr

3.6.1.6 Limitations and Opportunities

One of the defining features of the p(i|o) model is the relative scarcity of ignition data and the diversity of underlying causes and characteristics of related grid failure and outage events. This limits the approaches available to predict ignition rates and requires the introduction of model structure and other assumptions based on first principles. Future work can continue to refine the event groupings that provide shared response to weather and fuel conditions. Model structures capable of predicting well based upon limited training data should continue to be tested as possible alternative algorithms.

The ideal p(i|o) model would have perfect information about fault current, duration, and path to ground as well as finer resolution in time and space of fuel conditions, soil moisture, etc. While this data that is not currently sensed and available, knowledge about their potential influence can factor into model structure and assumptions as well as guide future data gathering initiatives.

4 Future Plans

Several future developments are planned for the distribution event probability models. *Figure 76* provides an overview of RaDA team development plans with Distribution Event Probability Model plans in highlighted in blue font.

Products	2026 WMP	Future WMP

Event Probability Models	• Data Quality Improvement
	• Model Refresh / Calibration
	• Asset Models – Distribution
	  • Switches
	  • Fuses
	  • Capacitor Banks
	  • Voltage Regulators
	• Asset Health – Tags
	• Work Plan Model Causality
	• Animal/Third Party – Distribution
	• Vegetation
	  • Tree Mortality
	  • Time Series Update
	• Insulator Contamination Update
		• Data Quality Improvements
		• Model Refresh / Calibration
		• Full Year Models
		• Asset Model – Distribution
		  • Secondary Conductor
		• Underground Assets
		• Lightning
		• Seismic – Distribution
		• Third Party – Transmission

Risk Models	• WDRM v4 • WTRM v2 • Public Safety Risk Model v2 • Reliability Risk Model v1 • Integrated Grid Planning	• WDRM Update • WTRM Update • Risk Mitigation Automation • Risk Reduction Reporting
Consequence Models	• Wildfire Consequence v4 • Wildfire Egress Impact • Wildfire Suppression Impact • Public Safety Consequence v2 • Reliability Consequence v1	• Wildfire Consequence Update • Conflagration • Vulnerable Communities • WUI Population Growth

Figure 76 - RaDA Product Plan – Distribution Event Probability Models

General development plans for all event probability of failure/outage models include:

- Develop suite of probability of outage given failure models to support reliability risk modeling.
- Improved situational awareness of model performance during wildfire season.
- Enhanced tooling to identify and explain changes between model releases.
- Enhanced tooling for automation and ML operations.

The Equipment Asset models will be expanded by developing:

- New Underground equipment models for:
  - Conductors
  - Switches
  - Transformers
- Secondary Conductor model will be transitioned from a spatial pixel model into an asset model.
- Improved equipment failure reporting in the Asset Failure Data Collection.

The Contact From Object models will be improved through:

- Migration to temporal use of multi-year satellite imagery.
- Potential use of inspections and tags as features, especially for the Vegetation models.
- Failure location data quality enhancements.

The Equipment Asset models will also undergo specific development for integration with Integrated Grid Planning (IGP):

- Enhance probability of failure models to predict for both wildfire season and full calendar year.
- Create a methodology to expand predictions from 1-year to 10-year failure probabilities.

Future Probability of Ignition model development will explore:

- Incorporation of new and improved FPI model predictions from Meteorology.
- Improvements to the model data pipeline, including a more flexible feature regularization approach.
- Additional data informing estimates of the effectiveness of the EPSS program.
- Exploratory work on imputing and incorporating fault characteristics into the model.