

*Southern California Edison*

***WSD-011 – Resolution implementing the requirements of Public Utilities Code Sections 8389(d)(1), (2) and (4) related to catastrophic wildfire caused by electrical corporations subject to the Commission’s regulatory authority***

**DATA REQUEST SET W S D - S C E - 0 0 4**

**To: WSD**

**Prepared by: Matthew Mendoza**

**Job Title: Predictive Analytics/Data Science, Advisor**

**Received Date: 3/12/2021**

**Response Date: 3/17/2021**

---

**Question 002:**

Regarding the dataset SCE used for the 2021 Wildfire Risk Model:

- 2(a). Confirm that SCE utilized outage data as a proxy for ignition data to train the model as stated at the 2/22/21 workshop. If this is not correct, explain what dataset was utilized.
- 2(b). How many years of outage data did SCE use to train the model?
- 2(c). How many years of outage data did SCE use to test to model?
- 2(d). How many events did SCE utilize to train the model?
- 2(e). How many events did SCE utilize to test the model?
- 2(f). For outage data used by SCE, was it possible for SCE to determine the exact location of the issue leading to an outage for each outage event? If it was not possible, did SCE utilize the outage event as part of the training dataset?

**Response to Question 002:**

- a) Not all outages can lead to sparks. For example, many outages are due to factors such as underground, source loss, and de-energization for upstream repair. Furthermore, not all spark-causing outages lead to ignitions as ignitions may depend on the presence of other factors such as weather and fuels. The probability of ignition model is therefore the combined probability of two outcomes: the likelihood that an outage leads to a spark and the likelihood that a spark leads to a fire. Outage data is used to model the likelihood that an outage at a specific piece of equipment will cause a spark, and historical fire data is used to model the frequency that a spark from a given piece of equipment leads to a fire. Saying “outage data is a proxy for ignition” is only encompassing half of the failure picture that is modeled.
- b) Ten years of outage data is used to train the model.
- c) A fraction of randomly sampled events from the ten-year data is held for testing. Furthermore, an additional year, distinct from the train/test decade, is withheld to validate the accuracy over time.
- d) Roughly 54,000 events are used to train the model.
- e) Roughly 23,000 events are used to test the model.
- f) The precise location is associated with the affected structure, which is identifiable for a majority of outages used in building the model. Outages without a location recorded and whose causes are those of interest for modeling are still included in the model building. The probability of failure is associated with each possible structure based on outage notes, nearest protection device, or circuit affected. No outage that could lead to a spark is omitted from the model building.