

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 M G R A - S C E - 0 0 6

To: MGRA

Prepared by: Matthew Mendoza

Job Title: Predictive Analytics/Data Science, Advisor

Received Date: 3/3/2021

Response Date: 3/8/2021

Question 003:

Describe the algorithm(s) used by the POI model. “Machine learning” is not sufficient detail.

Response to Question 003:

The POI model uses a form of ensemble model called a “Random Forest Classifier.” Ensemble models use a group or “ensemble” of decision trees to make classifications. The Random Forest Classifier takes subsets of random samples of the training data set and fits a decision tree on the independent variables that best discriminates between the classes of interest. The order with which the algorithm chooses variables to build out the decision trees is also randomized. Future data is classified based on the consensus of “votes” of the ensemble of decision trees.

Classifications in the context of POI models determine whether an asset experienced a failure given its physical features and the environmental and electrical stresses, and other data that the asset may experience. The Random Forest Classifier, therefore, creates a congress of logic trees that determine the likelihood of an asset failure.