**Config.json File Documentation**

**simulation\_metadata\_name**

* File name of the simulation metadata csv file
* Example value: "summary\_results\_algorithm\_Knoxville\_TN\_AMY\_brief\_2.csv"

**weight\_of\_baseline\_data**

* Since the simulation data is imbalanced, meaning that the non-fault data accounts for only a small portion of the whole dataset. This configuration is a multiplier to duplicate the current non-fault data records. Ideally, the fault and non-fault data sets should be given equivalent weights.
* Example value: 1

**random\_state**

* Random seed value for all random processes in the code. This creates replicable random forest results.
* Example value: 42 (this is a commonly used default, but any integer value will suffice)

**number of trees**

* Number of trees for random forest or gradient boosting algorithms
* Example value: 25 (25 is a default value; 100 would be more typical; sensitivity analysis is commonly conducted to identify the value beyond which performance fails to improve appreciably)

**considered\_sensors\_in\_sensor\_inaccuracy\_analysis**

* Subset of sensors (as a string list) to be considered in sensor inaccuracy analysis
* Example value: ["cooling\_electricity [W]", "electricity\_facility [W]", "whole\_building\_facility\_total\_hvac\_electric\_demand\_power [W]"]

**sensor\_category\_dict**

* A dictionary to map each sensor name to its sensor type. This is used in the sensor inaccuracy analysis
* Example value: {"cooling\_electricity [W]": "electricity\_meter",

"electricity\_facility [W]": "electricity\_meter", "whole\_building\_facility\_total\_hvac\_electric\_demand\_power [W]": "electricity\_meter"}

**sensor\_fault\_probability\_table\_dir**

* The csv file name of the sensor fault probability table. Explanation of the fault probability table can be found in “documentations/Fault\_Probability\_Table\_Description.docx”
* Example value: "sensor\_fault\_probability\_table.csv"

**selected\_fault\_types\_for\_sensor\_inaccuracy\_analysis**

* Subset of fault types (as a string list) to be considered for sensor inaccuracy analysis
* Example value: ["air\_handling\_unit\_fan\_motor\_degradation",

"biased\_economizer\_sensor\_mixed\_t",

"duct\_fouling"]

**sensor\_group\_info\_dir**

* The csv file name containing the sensor group information. The sensors are grouped as follows: basic, moderate, rich, and unlikely. Explanation of the sensor group information file can be found in “documentations/Sensor\_Group\_Info\_Description.docx”
* Example value: "Suggested Sensor Sets.csv"

**fault\_type\_list\_sensor\_cost\_analysis**

* A subset of fault types to be considered in the sensor cost analysis.
* Example value: ["thermostat\_bias", "economizer\_opening\_stuck"]

**baseline\_sensor\_set\_sensor\_cost\_analysis**

* Baseline sensor set (as a string list), which is used as an input for the sensor cost analysis. This could be an existing sensor set for an operational building or a base design set for a new project.
* Example value: [“room\_103\_zone\_mean\_air\_temperature [C]”,

“room\_104\_zone\_mean\_air\_temperature [C]”]

**candidate\_sensor\_set\_sensor\_cost\_analysis**

* Candidate (proposed) sensor set (as a string list), which is used for the settings in sensor cost analysis. It defines the candidate sensors to be evaluated against the baseline set. In many cases, this set would include most or all of the sensors in the baseline set plus some additional sensors to improve FDD performance.
* Example value: ["room\_204\_zone\_air\_relative\_humidity [%]", "electricity\_facility [W]"]

**fault\_prevalence\_dict**

* A dictionary with fault types as ‘keys’ and the fault prevalence fractions as ‘values’. It is used to aggregate sensor opportunity cost across fault types.
* Example value: {"air\_handling\_unit\_fan\_motor\_degradation": 0.255,

"biased\_economizer\_sensor\_mixed\_t": 0.19}

**guidance\_bldg\_dict**

* A dictionary with the building name as a ‘key’. The ‘value’ is not used and therefore assigned a dummy value. This dictionary only applied to the ‘general\_guidance’ case and the provided example value is currently the only valid value.
* Example value: {"small\_commercial\_building": "None"}

**algorithm\_dict**

* A dictionary with supported machine learning algorithms as ‘keys’. The ‘values’ are not used and therefore set to dummy values.
* Example value: {"random\_forest": "None", "gradient\_boosting": "None"}