**Results:**

The purpose of this work was to answer five specific questions regarding data and attributes of smart homes. These five questions are as follows:

1. Does a home use electricity for heat?
2. Does a home have a cooling system?
3. Does a home use electricity for heating water?
4. Does a home have solar panels?
5. Does a home use electricity for stoves?

To answer these questions, machine learning models were built in an attempt to classify this information accordingly. To be more precise, the models were built in Python using different APIs to facilitate the model-building process. TensorFlow, an end-to-end open-source platform for machine learning was chosen as the main software library to use for its ease of access to tools for training and inference of deep neural networks. To aid in the output and interpretation of results, the scikit-learn API was used for predictive data analysis, along with imbalanced-learn which was used to deal with certain datasets in this study which had largely imbalanced classes.

While each question was answered with its own respective machine learning model, certain general steps were taken for all five models. To begin, all models shared the same features. These features were the monthly totals of electricity consumption for each building respectively. This data was read from a csv file and stored in a data frame via Pandas. In a similar fashion, the column of data respective to the question being answered by the model was also stored into a Pandas data frame. The approach for all five of the models in this study was that of binary classification. All values in the labels data frame were converted to either 1 or 0, representative of whether or not an individual building possessed the quality being searched for by the question posed. In regard to preprocessing the data, all features for each model were transformed with sklearn’s StandardScalar(). This was performed due to the wide ranges of values in the features, which can cause the machine learning model to behave poorly as data is not distributed normally. To adjust the data accordingly, the function performs standardization on the dataset following the equation below.

Diagram, text

Description automatically generated with medium confidence

*Equation 1: Standardization*

The function also adjusts the mean of individual features to zero, as well as the standard deviation of individual features to one, so that the data used in the machine learning model will have a common scale.

A picture containing text, clock

Description automatically generated

*Equation 2: Standard Deviation*

A picture containing text

Description automatically generated

*Equation 3: Mean*