

Smart Home – Temperature Prediction

1.INTRODUCTION

A smart home's devices are connected with each other and can be accessed through one central point like a smartphone, tablet, laptop, or game console. Door locks, televisions, thermostats, home monitors, cameras, lights, and even appliances such as the refrigerator can be controlled through one home automation system. Smart Wi-Fi thermostats have moved well beyond the function they were originally designed for controlling heating and cooling comfort in buildings. They are now also learning from occupant behaviours and permit occupants to control their comfort remotely. Thermal comfort in buildings has been managed for many years by thermostats. At a most basic level, a thermostat allows a resident to set a desired indoor temperature, a means to sense actual temperature within the thermostat housing, and a means to signal the heating and/or cooling devices to turn on or off in order to affect control of the heating, ventilating, and air conditioning (HVAC) system in order to equilibrate the room temperature to the set point temperature. Thermostats use sensors such as thermistors or thermal diodes to measure temperature, they also often include humidity sensors for measuring humidity and microprocessor-based circuitry to control the HVAC system and operate based upon user-defined set point schedules. This project seeks to go beyond this state of the art by utilizing smart Wi-Fi thermostat data in residences to develop dynamic predictive models for room temperature and cooling/heating demand. While efforts are being made around the world to minimize greenhouse gas emissions and make progress towards a more sustainable society, global energy demand continues to rise. Building energy consumption accounts for 20–40% of the total global energy consumption and Heating, Ventilation, Air Conditioning (HVAC) answer for around 50% of this amount. Therefore, implementing energy efficiency-related strategies and optimization techniques in buildings is a critical step in reducing global energy consumption.

In this project we will just take the data that is generated by the sensors by The University of CEU Cardenal Herrera (CEU-UCH)-Spain. We will preprocess the data and pass it to the Regression algorithms such as Linear Regression, Random forest, LightGDBM, and Xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pickle format. We will be doing flask integration and IBM

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

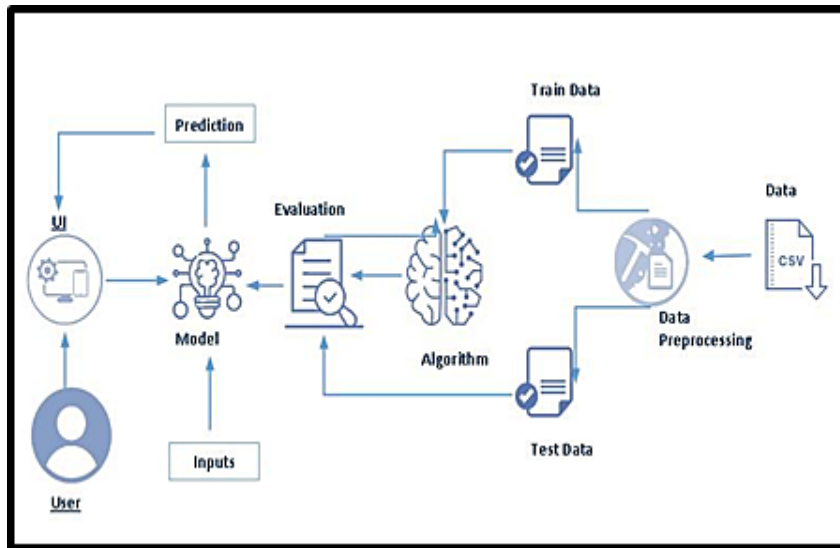
The existing temperature monitoring systems generally use weather stations that use multiple instruments such as thermometers, etc. to measure temperature changes. Most of these instruments use simple analog technology which is later physically recorded and stored in a data base. Existing temperature monitoring systems that are used in the field generally consist of unconventional and heavy machinery that consists of numerous moving parts that require constant maintenance and need to be manually monitored and changed frequently. The use of thermometers to measure external temperature; however accurate is still outdated and constantly needs to be manually checked for any change in temperature. Data that is collected by the instruments needs to be manually transferred from the logger to a laptop or computer via a cable. Existing systems consist of large and heavy instruments that occupy a lot of space hence making it difficult to install them in remote location and places which have limited space. The instruments used in the existing systems are expensive and add up to the already high cost of installation and maintenance.

2.2 Proposed Solution

The system proposed is an advanced solution for weather monitoring that uses IoT to make its real time data easily accessible over a very wide range. The system deals with monitoring temperature. Our proposed 'Smart home temperature prediction system' unlike conventional temperature monitoring instruments is very small and compact allowing it to be installed easily on smart home. It is light and portable; this advantage allows us to easily carry it to remote location for installation. The power requirements for our system (sensors and boards) is much less compared to the existing instruments in the market hence enabling us to use solar cells as power supply. This not only cuts down on cost but allows us to leave the monitoring system in remote, areas where power is not easily available, for long periods of time. Addition of solar panels also helps our design be eco-friendly. The sensors used in our product are much cheaper compared to the ones that are used in the existing weather monitoring systems making our design more cost effective. These sensors send the data to a web page and the sensor data is plotted as graphical statistics. The data uploaded to the web page can easily be accessible from anywhere in the world. The data gathered in these web pages can also be used for future references. Unlike the existing system where data has to be physically transferred. Due to the presence of fewer moving parts less amount of maintenance will be needed cutting down on maintenance charges. The product even consists of an app that sends notifications as an effective alert system to warn people about sudden and drastic temperature changes.

3.THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware/Software designing

To complete this project, you must required following software's, concepts and packages

1. **Anaconda navigator and pharm:**
 - a. download anaconda navigator
2. **Python packages:**
 - Open anaconda prompt as administrator
 - Type "pip install numpy" and click enter.
 - Type "pip install pandas" and click enter.
 - Type "pip install scikit-learn" and click enter.
 - Type "pip install matplotlib" and click enter.
 - Type "pip install scipy" and click enter.
 - Type "pip install pickle-mixin" and click enter.
 - Type "pip install seaborn" and click enter.
 - Type "pip install Flask" and click enter.

4.EXPERIMENTAL INVESTIGATIONS

1. Know fundamental concepts and techniques used for machine learning.
2. Gain a broad understanding about data.
3. Have knowledge on pre-processing the data/transformation techniques on outlier and some visualization concepts.

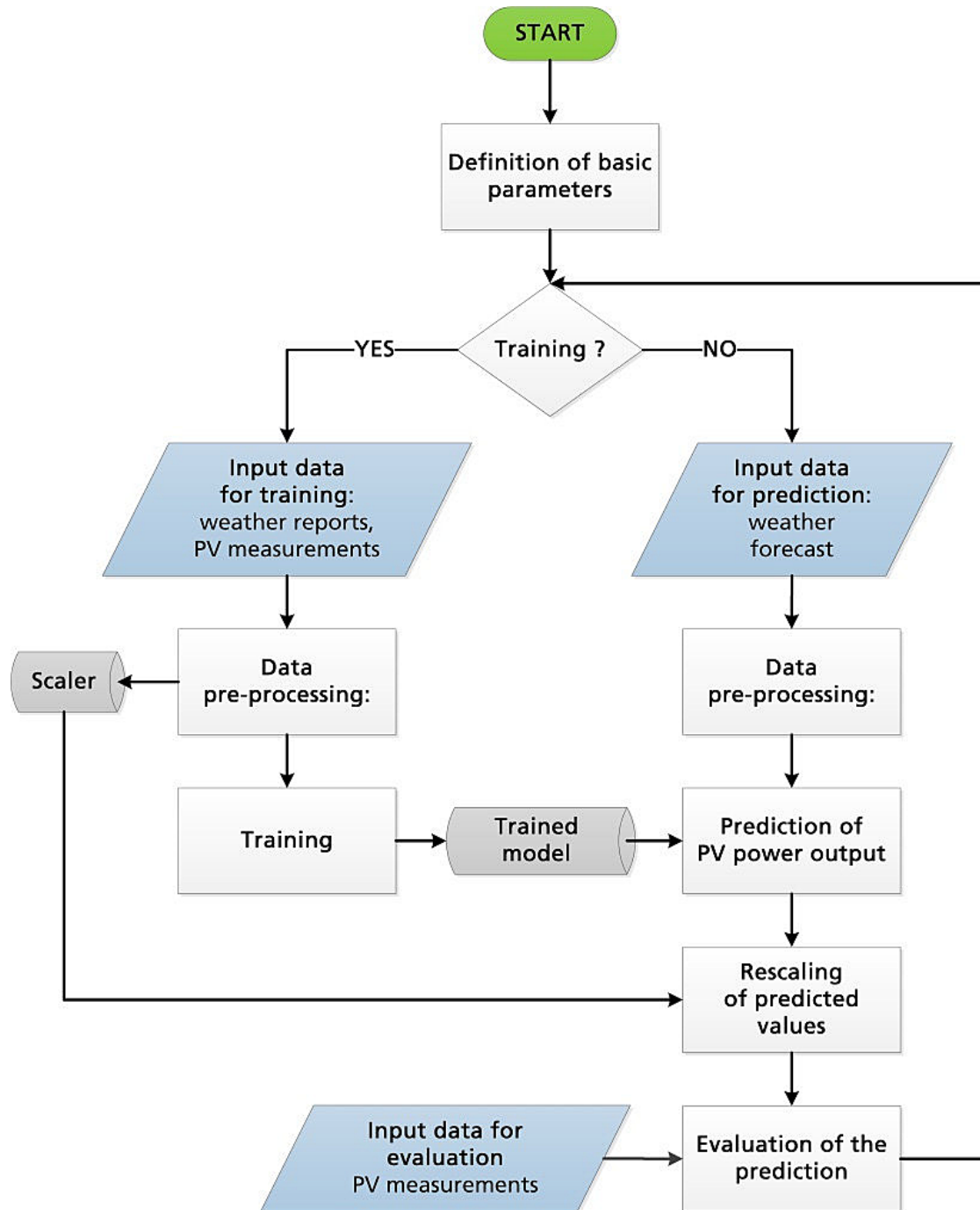
Project Flow:

1. User interacts with the UI to enter the input.
2. Entered input is analyzed by the model which is integrated.
3. Once model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities listed below,

1. Data collection
 - a. Collect the dataset or create the dataset
2. Visualizing and analyzing data
 - Univariate analysis
 - Bivariate analysis
 - Multivariate analysis
 - Descriptive analysis
3. Data pre-processing
 - a. Checking for null values
 - b. Handling outlier
 - c. Handling categorical data
 - d. Splitting data into train and test

5. FLOWCHART



6.RESULTS

Activity 1: Importing the libraries

Import the necessary libraries as shown in the image.

```
: #importing the libraries
import pandas as pd
import numpy as np
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import RandomForestRegressor
import xgboost as xgb
import lightgbm as lgb
```

Activity 2: Read the Dataset

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called `read_csv()` to read the dataset. As a parameter we have to give the directory of csv file.

```
#load the dataset
df = pd.read_csv(r"C:\Users\HP\Downloads\Temperature\Regression\Dataset\data.csv")
```

```
#displaying the first five rows
df.head()
```

	Date	Time	CO2_(dinning-room)	CO2_room	Relative_humidity_(dinning-room)	Relative_humidity_room	Lighting_(dinning-room)	Lighting_room	Meteo_Rain	Meteo_Sun_dusk
0	13-03-12	11:45	216.500	221.920	39.9125	42.4150	81.6650	113.520	0.0	623.360
1	13-03-12	12:00	219.947	220.363	39.9267	42.2453	81.7413	113.605	0.0	623.211
2	13-03-12	12:15	219.403	218.933	39.7720	42.2267	81.4240	113.600	0.0	622.656
3	13-03-12	12:30	218.613	217.045	39.7760	42.0987	81.5013	113.344	0.0	622.571
4	13-03-12	12:45	217.714	216.080	39.7757	42.0586	81.4657	113.034	0.0	622.400

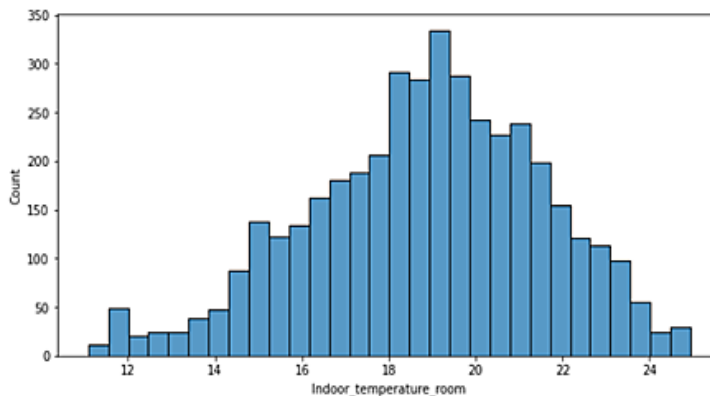
Activity 3: Univariate analysis

In simple words, univariate analysis is understanding the data with single feature. Here we have displayed two different graphs such as distplot and countplot.

Seaborn package provides a wonderful function distplot. With the help of distplot, we can find the distribution of the feature. To make multiple graphs in a single plot, we use subplot.

univariate analysis

```
plt.figure(figsize=(10,5))
sns.histplot(data = df, x='Indoor_temperature_room',)
<AxesSubplot:xlabel='Indoor_temperature_room', ylabel='Count'>
```



1. From the plot we came to know, Indoor_temperature_room column, which is our output column it follows normal distribution.

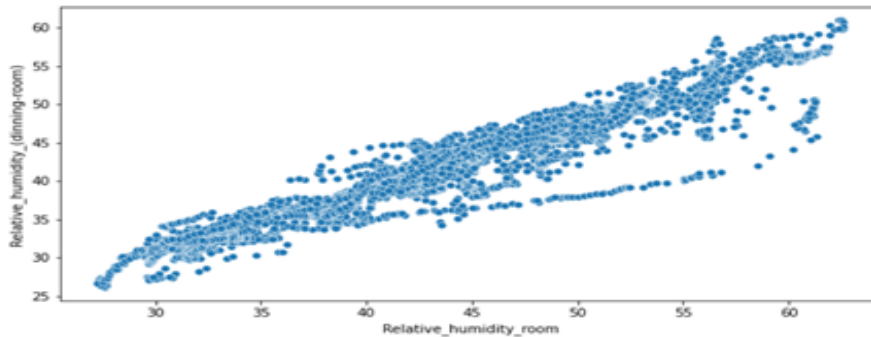
Activity 4: Bivariate analysis

Scatter Plot():

Scatter plots are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a Cartesian system. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis.

bi variate analysis

```
plt.figure(figsize =(10,5))
sns.scatterplot(data = df, x = 'Relative_humidity_room', y = 'Relative_humidity_(dinning-room)')
<AxesSubplot:xlabel='Relative_humidity_room', ylabel='Relative_humidity_(dinning-room)'
```

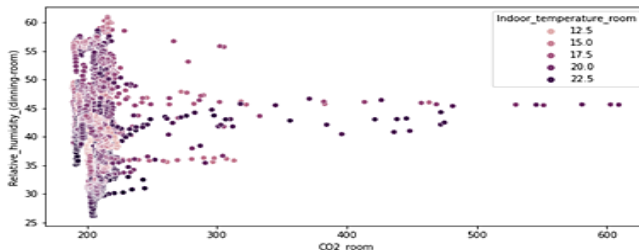


Activity 5: Multivariate analysis

In simple words, multivariate analysis is to find the relation between multiple features. Here we have used swarm plot from Seaborn package.

multivariate analysis

```
plt.figure(figsize =(10,5))
sns.scatterplot(data = df, x = 'CO2_room', y = 'Relative_humidity_(dinning-room)', hue='Indoor_temperature_room')
<AxesSubplot:xlabel='CO2_room', ylabel='Relative_humidity_(dinning-room)'
```



Activity 6: Descriptive analysis

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.


```
df.describe()
```

	CO2_(dinning-room)	CO2_room	Relative_humidity_(dinning-room)	Relative_humidity_room	Lighting_(dinning-room)	Lighting_room	Meteo_Rain	Meteo_Sun_dusk	Meteo_Sun_dawn
count	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000
mean	205.599835	209.611623	42.389879	44.546099	28.970248	42.335496	0.038756	335.094312	1.000000
std	22.763114	24.183477	7.215405	8.297436	25.684356	42.602571	0.187128	304.513038	1.000000
min	187.339000	188.907000	26.173300	27.256000	10.740000	11.328000	0.000000	0.606667	0.000000
25%	200.228000	201.707000	36.088000	38.446700	11.540700	13.509300	0.000000	0.650000	0.000000
50%	205.131000	208.907000	42.776000	44.802700	14.126700	22.085300	0.000000	612.821000	0.000000
75%	210.616000	212.331000	47.584000	50.301300	40.034700	55.094000	0.000000	619.712000	2.000000
max	594.389000	609.237000	60.957300	62.594700	111.797000	162.965000	1.000000	625.003000	6.000000

Milestone 3: Data Pre-processing

As we have understood how the data is lets pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much of randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

2. Handling missing values
3. Handling categorical data
4. Handling outliers
5. Scaling Techniques
6. Splitting dataset into training and test set

Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

Activity 1: Checking for null values

7. Let's find the shape of our dataset first, to find the shape of our data, df.shape method is used. To find the data type, df.info() function is used.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4137 entries, 0 to 4136
Data columns (total 18 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0    Date                                     4137 non-null   object
1    Time                                     4137 non-null   object
2    CO2_(dinning-room)                     4137 non-null   float64
3    CO2_room                                4137 non-null   float64
4    Relative_humidity_(dinning-room)       4137 non-null   float64
5    Relative_humidity_room                 4137 non-null   float64
6    Lighting_(dinning-room)                4137 non-null   float64
7    Lighting_room                          4137 non-null   float64
8    Meteo_Rain                             4137 non-null   float64
9    Meteo_Sun_dusk                         4137 non-null   float64
10   Meteo_Wind                             4137 non-null   float64
11   Meteo_Sun_light_in_west_facade          4137 non-null   float64
12   Meteo_Sun_light_in_east_facade          4137 non-null   float64
13   Meteo_Sun_light_in_south_facade         4137 non-null   float64
14   Meteo_Sun_irradiance                    4137 non-null   float64
15   Outdoor_relative_humidity_Sensor        4137 non-null   float64
16   Day_of_The_week                         4137 non-null   float64
17   Indoor_temperature_room                 4137 non-null   float64
dtypes: float64(16), object(2)
memory usage: 581.9+ KB
```

8.

9. For checking the null values, `df.isnull()` function is used. To sum those null values we use `.sum()` function to it. From the below image we found that there are no null values present in our dataset. So we can skip handling of missing values step.

From the above code of analysis, we can infer that columns Do not have any Null Values, so we don't perform null values operations on this dataset.

Activity 2: Handling Categorical Values

As we can see our dataset has categorical data we must convert the categorical data to integer encoding or binary encoding.

To convert the categorical features into numerical features we use encoding techniques. There are several techniques but in our project we are using manual encoding with the help of list comprehension.

10. In our dataset, we don't have any categorical values and most of the values we have are float so, we don't perform any encoding techniques.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4137 entries, 0 to 4136
Data columns (total 18 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0    Date                                     4137 non-null   object
1    Time                                     4137 non-null   object
2    CO2_(dinning-room)                     4137 non-null   float64
3    CO2_room                                4137 non-null   float64
4    Relative_humidity_(dinning-room)       4137 non-null   float64
5    Relative_humidity_room                 4137 non-null   float64
6    Lighting_(dinning-room)                4137 non-null   float64
7    Lighting_room                          4137 non-null   float64
8    Meteo_Rain                             4137 non-null   float64
9    Meteo_Sun_dusk                         4137 non-null   float64
10   Meteo_Wind                             4137 non-null   float64
11   Meteo_Sun_light_in_west_facade          4137 non-null   float64
12   Meteo_Sun_light_in_east_facade          4137 non-null   float64
13   Meteo_Sun_light_in_south_facade         4137 non-null   float64
14   Meteo_Sun_irradiance                    4137 non-null   float64
15   Outdoor_relative_humidity_Sensor        4137 non-null   float64
16   Day_of_The_week                         4137 non-null   float64
17   Indoor_temperature_room                 4137 non-null   float64
dtypes: float64(16), object(2)
memory usage: 581.9+ KB
```

11.

Activity 3: Scaling the Data

Scaling is one the important process, we have to perform on the dataset, because of data measures in different ranges can leads to mislead in prediction

Models such as linear regression need scaled data, as they follow distance based method and Gradient Descent and Tree concept no need of scaling.

```
: from sklearn.preprocessing import StandardScaler  
sc = StandardScaler()  
x_train_scaled = sc.fit_transform(x_train)  
x_test_scaled = sc.transform(x_test)
```

We will only perform scaling on the input values and in this project scaling is performed for only linear regression

Activity 4: Splitting data into train and test

Now let's split the Dataset into train and test sets

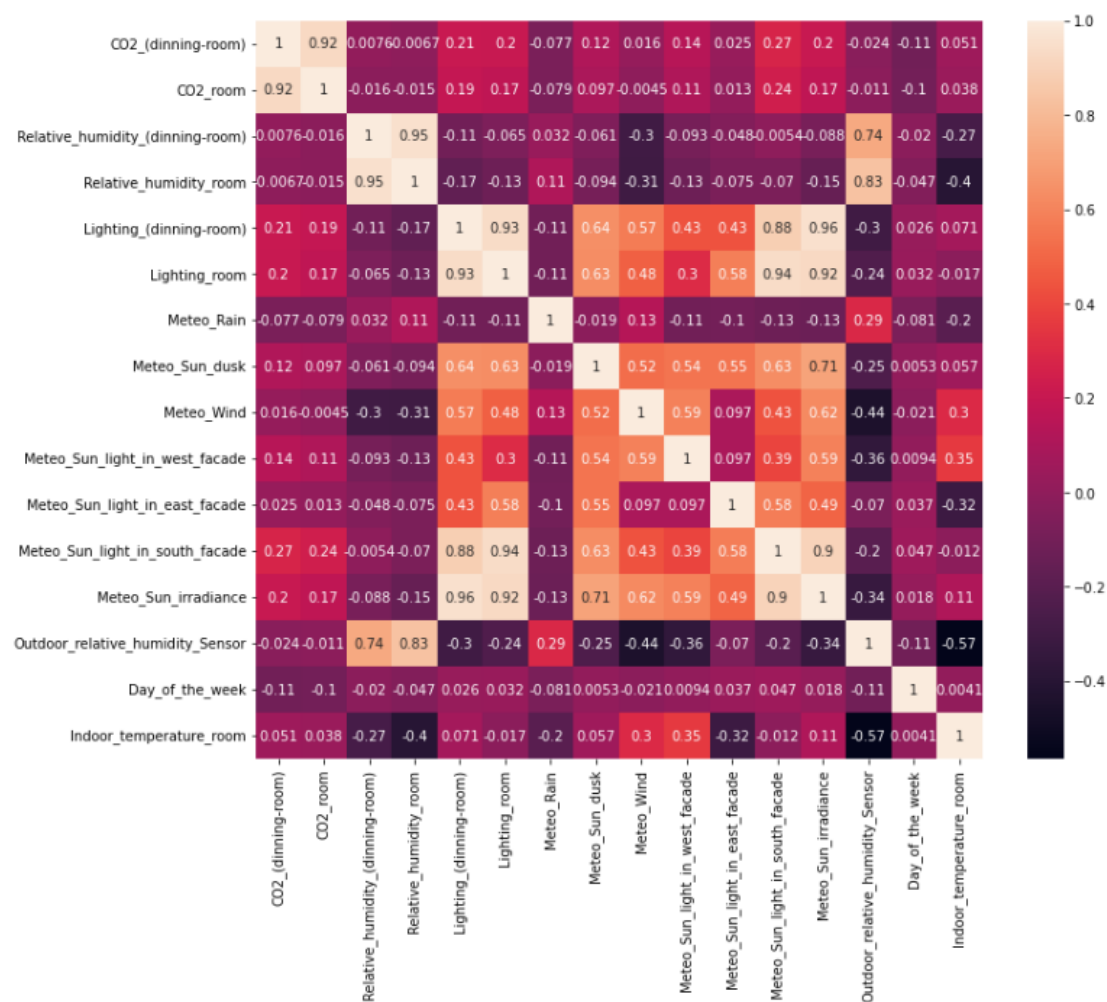
Changes: first split the dataset into x and y and then split the data set

Here x and y variables are created. On x variable, df is passed with dropping the target variable.

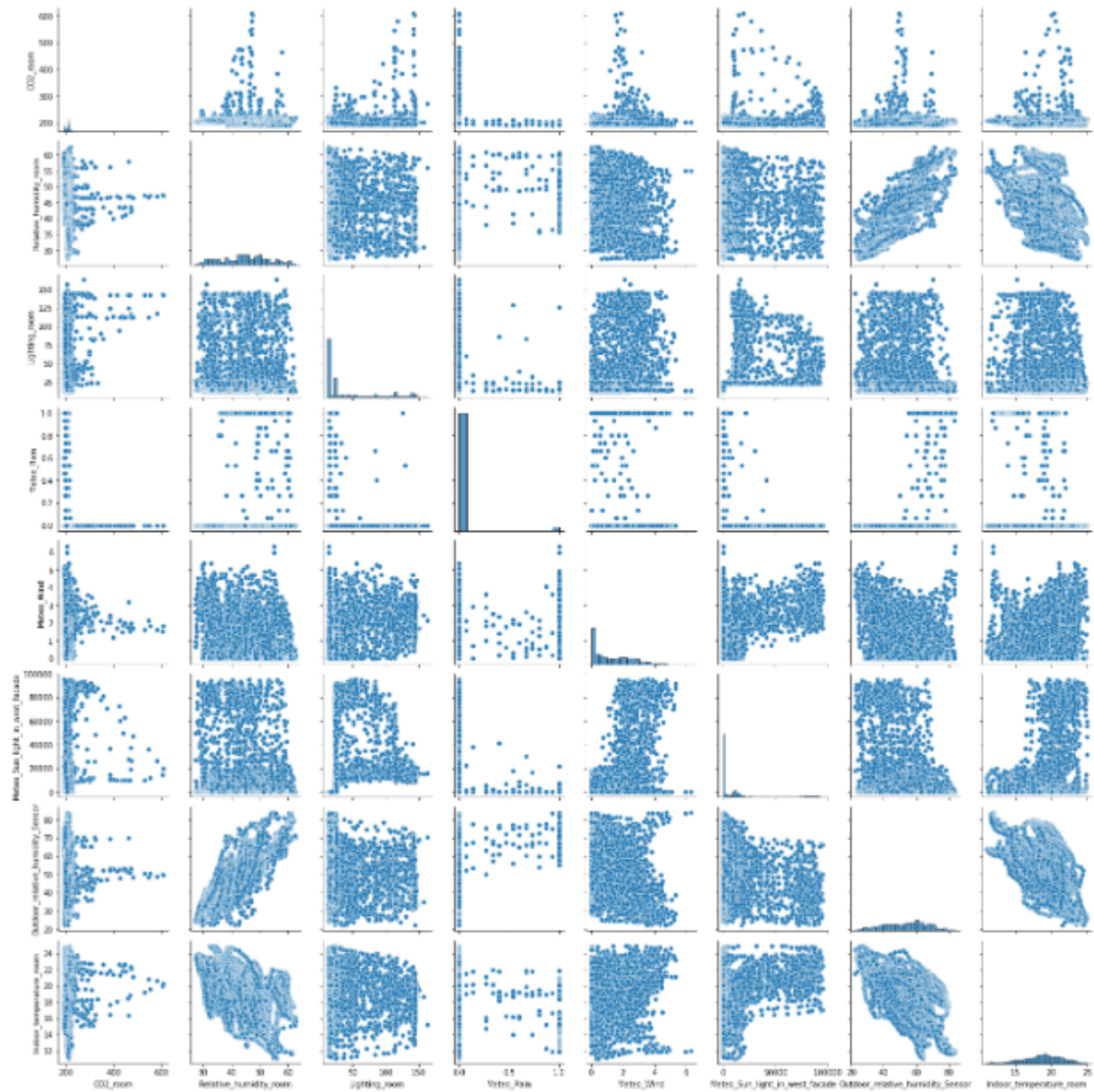
```
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3,random_state=1)
```

And on y target variable is passed. For splitting training and testing data we are using train_test_split() function from sklearn. As parameters, we are passing x, y, test_size, random_state.

Heatmap



Pairplot



Milestone 4: Model Building

Now our data is cleaned and it's time to build the model. We can train our data on different algorithms. For this project we are applying four classification algorithms. The best model is saved based on its performance.

Activity 1: Liner Regression model

A function named Linear Regression is created and train and test data are passed as the parameters. Inside the function, Linear Regression algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, used r2 score

```
from sklearn.linear_model import LinearRegression
lir = LinearRegression()
lir.fit(x_train_scaled,y_train)

LinearRegression()

pred = lir.predict(x_test_scaled)

r2_score(pred,y_test)
-0.44264951676880626
```

Activity 2: Random forest model

A function named randomForest is created and train and test data are passed as the parameters. Inside the function, RandomForestRegressor algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluation

```

rf = RandomForestRegressor()

rf.fit(x_train,y_train)
RandomForestRegressor()

pred = rf.predict(x_test)

pred
array([23.24284434, 17.729198 , 21.2047528 , ..., 20.21012992,
       17.78471072, 18.859208  ])

from sklearn.metrics import r2_score
r2_score(y_test,pred)
0.8705801769012985

```

ng

the model, used r2 score

Activity 3: Light Gradient Boost model

A function named lg is created and train and test data are passed as the parameters. Inside the function, LGBM Regressor algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model used r2 score.

```

lg = lgb.LGBMRegressor()

lg.fit(x_train,y_train)
LGBMRegressor()

pred = lg.predict(x_test)

r2_score(y_test,pred)
0.8569554082913747

```

Activity 4: Xgboost model

A function named xgboost is created and train and test data are passed as the parameters. Inside the function, GradientBoostingClassifier algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new

variable. For evaluating the model used r2score

Now let's see the performance of all the models and save the best model

```
xg = xgb.XGBRegressor()

xg.fit(x_train,y_train)

XGBRegressor(base_score=0.5, booster='gbtree', callbacks=None,
              colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
              early_stopping_rounds=None, enable_categorical=False,
              eval_metric=None, gamma=0, gpu_id=-1, grow_policy='depthwise',
              importance_type=None, interaction_constraints='',
              learning_rate=0.300000012, max_bin=256, max_cat_to_onehot=4,
              max_delta_step=0, max_depth=6, max_leaves=0, min_child_weight=1,
              missing=nan, monotone_constraints=(), n_estimators=100, n_jobs=0,
              num_parallel_tree=1, predictor='auto', random_state=0, reg_alpha=0,
              reg_lambda=1, ...)

pred = xg.predict(x_test)

r2_score(y_test,pred)

0.8589766550598491
```

Activity 5: Evaluating performance of the model and saving the model

From sklearn, cross_val_score is used to evaluate the score of the model. On the parameters, we have given rf (model name). Our model is performing well. So, we are saving the model by pickle.dump().

```
import pickle
pickle.dump(rf,open('temperature.pkl','wb'))
```

Milestone 5: Application Building

In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks

12. Building HTML Pages

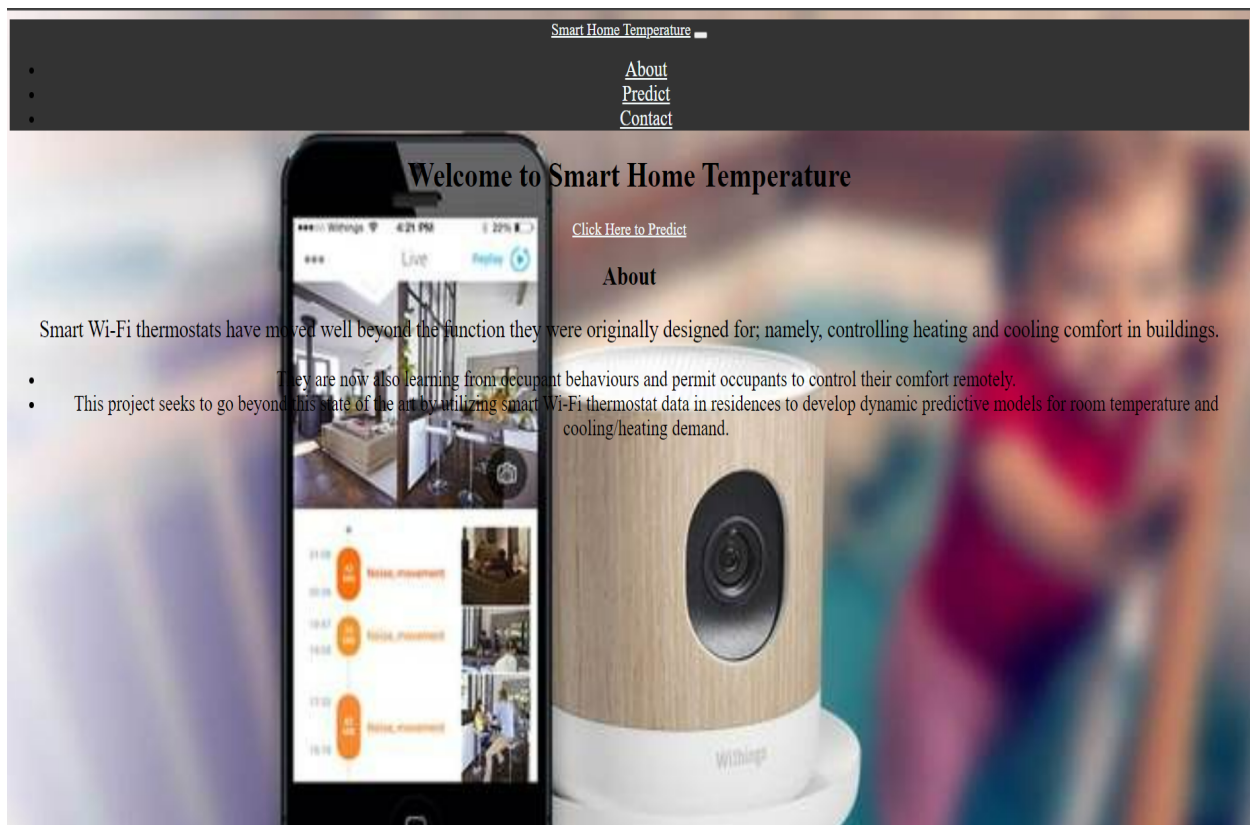
13. Building server side script

Activity1: Building Html Pages:

For this project create three HTML files namely

1. index.html
2. predict.html and save them in templates folder.

Let's see how our index.html page looks like:



Welcome to Smart Home Temperature

CO2: CO2 in the room

Humidity: Humidity in the room

Lighting: Lighting in the room

Rain: Rain in Last 15 Minutes

Wind: Wind at outside

Sunlight at West: Sunlight at west side of h

Outdoor Humidity: Humidity at outside

Submit

Your Room temperature will be °C

Smart Home Temperature

About

Contact

Welcome to Smart Home Temperature

CO2:

CO2 in the room

Humidity:

Humidity in the room

Lighting:

Lighting in the room

Rain:

Rain in Last 15 Minutes

Wind:

Wind at outside

Sunlight at West:

Sunlight at west side of h

Outdoor Humidity:

Humidity at outside

Submit

Your Room temperature will be 21.506377263900013 °C

9.CONCLUSION

To implement this need to deploy the sensor devices in the home for collecting the data and analysis. By deploying sensor devices in the environment, it will record real time data. It can cooperate with other objects through the network. Then the collected data and analysis results will be available to the end user through the Wi-Fi. The smart way to monitor home and an efficient, low-cost entrenched system is presented with different models in this paper.

APPENDIX

```
from flask import Flask, request, render_template
import pickle
import numpy as np
import pandas as pd

model = pickle.load(open(r'C:\Users\Rahul PK\Smart Home - Temperature
Prediction\Flask\temperature.pkl', 'rb'))

app = Flask(__name__)

@app.route("/")
def home():
    return render_template("index.html")

@app.route("/predict")
def predict():
    return render_template("predict.html")

@app.route('/output', methods = ['post', 'get'])
def output():
    # reading the inputs given by the user
    input_feature= [float(x) for x in request.form.values()]
    input_feature=np.array(input_feature)
    print(input_feature)

    names = ['CO2_room', 'Relative_humidity_room', 'Lighting_room',
'Meteo_Rain', 'Meteo_Wind', 'Meteo_Sun_light_in_west_facade',
'Outdoor_relative_humidity_Sensor']

    print(names)
```

```
data = pd.DataFrame(input_feature, columns=names)
print(data)
prediction=model.predict(data)
print(prediction)
return render_template('predict.html', prediction=prediction[0])

if __name__ == '__main__':
    app.run(debug = True)
```