**Capstone Project -1**

**Appliances Energy Prediction**

Business Problem Description – Dataset contains the house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. As per the description on UCI website, each wireless node transmitted the temperature and humidity conditions around 3.3 min, Then, the wireless data was averaged for 10 minutes periods. The energy data was logged every 10 minutes with m-bus energy meters. Combining this data with the weather data based on the date time columns

To find the key feature from the dataset which contributes to the most and the best prediction model for future prediction of the appliance Energy.

**Dataset Details –**

There are 19735 and 29 attributes. Key features from the dataset are

|  |  |
| --- | --- |
| **Columns** | **Description** |
| date time | year-month-day hour:minute:second |
| Appliances | energy use in Wh |
| lights | energy use of light fixtures in the house in Wh |
| T1 | Temperature in kitchen area in Celsius |
| RH\_1 | Humidity in kitchen area in % |
| T2 | Temperature in living room area in Celsius |
| RH\_2 | Humidity in living room area in % |
| T3 | Temperature in laundry room area |
| RH\_3 | Humidity in laundry room area in % |
| T4 | Temperature in office room in Celsius |
| RH\_4 | Humidity in office room in % |
| T5 | Temperature in bathroom in Celsius |
| RH\_5 | Humidity in bathroom in % |
| T6 | Temperature outside the building (north side) in Celsius |
| RH\_6 | Humidity outside the building (north side) in % |
| T7 | Temperature in ironing room in Celsius |
| RH\_7 | Humidity in ironing room in % |
| T8 | Temperature in teenager room 2 in Celsius |
| RH\_8 | Humidity in teenager room 2 in % |
| T9 | Temperature in parents room in Celsius |
| RH\_9 | Humidity in parents room in % |
| To | Temperature outside (from Chievres weather station) in Celsius |
| Pressure (from Chievres weather station) | in mm Hg |
| RH\_out | Humidity outside (from Chievres weather station) in % |
| Wind speed (from Chievres weather station) | in m/s |
| Visibility (from Chievres weather station) | in km |
| Tdewpoint (from Chievres weather station) | Â°C |
| rv1 | Random variable 1, nondimensional |
| rv2 | Random variable 2, nondimensional |

Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru). There are 2 random variables which has be explored more.

**Approach –**

1. **Data Preprocessing & Data Visualization –**
   * 1. Missing values Analysis and decision on whether to replace the missing values or delete the records.
     2. Boxplot, Histogram shows the distribution of data of different variables -using the libraries from Matplotlib and Seaborn
     3. Correlation plot informs about the relation between variables and analysis the correlation of all the features in the dataset.
     4. Feature selection – which columns are contributing as important features for the prediction.
2. **Data Exploration** – Analyze and plot the categorical and continuous feature summaries to see which feature is closely related with target variable - Appliance. This help us with deciding which feature are influencing the prediction.
3. **Data Cleaning and Feature engineering** – Encoding and imputing missing values in the data and checking for outliers with replacing with mean values and relabeling the values in categorical columns as to bring consistencies. Also, added additional columns for effective feature engineering.
4. **Model Experiments –** 
   1. Create a baseline model and compare with other models with key important features like Linear Regression, Ridge, Lasso, Support Vector Regressor, Random Forest, Gradient Boosting Regressor
   2. Model tuning – Using the GridsearchCV and RandomizedSearchCV, tune to model and evaluate the model accuracy by calculating the root mean square error.
   3. Model Evaluation- Comparing the results from above steps, we can determine the best model can used implemented for Appliance energy prediction.