**Appliance Energy Prediction**

**Sandeep R**

**Data science trainee,**

**AlmaBetter, Bangalore**

**Abstract:**

Electrical Appliance Require Power for their continuous operation, The power of an electrical appliances tells us how much electrical energy it transfers in a second. Power **P** is measured in watts **(W)** where **1W = 1 J/s** (joule/second). The amount of energy transferred from the mains appliance depends on the power rating of the appliance and the time for which it is switched on.

In this project I will be building a model which predicts the consumption of power measured in watts, depending on various factors inside a home such as temperature, humidity, and weather condition inside and outside home.

***Keywords: machine learning, weather, humidity, temperature.***

**1.Problem Statement**

The Dataset contains various features and Records for an interval of 10 mins which have been averaged. Consumption of Energy is recorded depending on various weather condition, temperature, and humidity around the house the task is to predict the consumption of energy by an appliance given weather, temperature and humidity features. In this dataset dependent feature will be Appliance and Remaining all the features will be independent features.

Independent Variables:

* date time year-month-day hour: minute: second
* T1, Temperature in kitchen area, in Celsius.
* T2, Temperature in living room area, in Celsius.
* T3, Temperature in laundry room area.
* T4, Temperature in office room, in Celsius.
* T5, Temperature in bathroom, in Celsius.
* T6, Temperature outside the building (north side), in Celsius
* T7, Temperature in ironing room, in Celsius
* T8, Temperature in teenager room 2, in Celsius
* T9, Temperature in parents’ room, in Celsius
* T\_out, Temperature outside (from  Chèvres weather station).
* RH1, Humidity in kitchen area, in %
* RH2, Humidity in living room area, in %
* RH3, Humidity in laundry room area, in %
* RH4, Humidity in office room, in %
* RH5, Humidity in bathroom, in %
* RH6, Humidity outside the building (north side), in %
* RH7, Humidity in ironing room, in %
* RH8, Humidity in teenager room 2, in %
* RH9, Humidity in parents’ room, in %
* RHout, Humidity outside (from Chievres weather station), in %
* Wind speed: Wind Speed (from  Chèvres weather station), in m/s
* Visibility: Visibility (from Chèvres weather station), in km
* Tdewpoint: the atmospheric temperature (varying according to pressure and humidity) below which water droplets begin to condense and dew can form

**2. Introduction**

The data set is at 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. 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. Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru) and merged with the experimental data sets using the date and time column

## **3. Energy Consumption** Electric energy consumption is the form of [energy consumption](https://en.wikipedia.org/wiki/Energy_consumption) that uses [electric energy](https://en.wikipedia.org/wiki/Electric_energy). Electric energy consumption is the actual energy demand made on existing electricity supply.

Electric and electronic devices consume electric energy to generate desired output (i.e., light, heat, motion, etc.). During operation, some part of the energy depending on the [electrical efficiency](https://en.wikipedia.org/wiki/Electrical_efficiency) is consumed in unintended output, such as waste heat

## **4. Finding wattage of Appliance****.**

 There are two ways to find the wattage an appliance uses:

* Stamped on the appliance

The wattage of most appliances is usually stamped on the bottom or back of the appliance, or on its nameplate. The wattage listed is the maximum power drawn by the appliance.

* Multiply the appliance ampere usage by the appliance voltage usage

If the wattage is not listed on the appliance, you can still estimate it by finding the electrical current draw (in amperes) and multiplying that by the voltage used by the appliance.

**5. Steps involved:**

* **Exploratory Data Analysis**

After loading the dataset we perform this method by comparing our target variable which is Appliances, with other independent variables. This process help us figuring out various aspects and relationships among the target and the independent variables. It gives us a better idea of which feature behaves in which manner compared to the target variable.

* **Null values Treatment**

Our dataset doesn’t contain any null values hence null value treatment is not required, I can continue to explore the dataset and work on the dataset to visualize and train the model.

* **Feature Selection**

All the features are used in this dataset. Mainly there are three different types of features namely, Weather type, temperature and humidity type of features. There are a total of 25 features of out of which 1 of the feature [Appliances] is target feature while the other features are independent features.

* **Standardization of features**

Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.

The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

* **Fitting different models**

For modelling I tried various Regression algorithms like:

1. **Linear Regression**
2. **Ridge Regression**
3. **Lasso Regression**
4. **Decision Tree Regressor**
5. **Support Vector Regression**
6. **Random Forest Regressor**

* **Tuning the hyperparameters for better accuracy**

Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting in case of tree based models

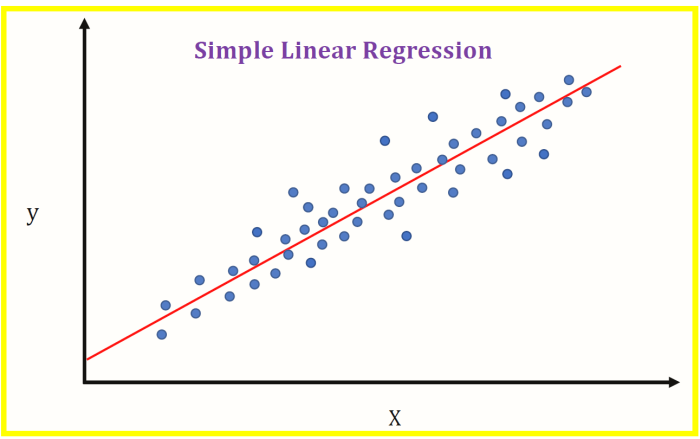
like Random Forest Regressor.

**6.1. Algorithms:**

1. **Linear Regression:**

Linear Regression is perhaps one of the most well-known and well understood algorithms in statistics and machine learning. The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

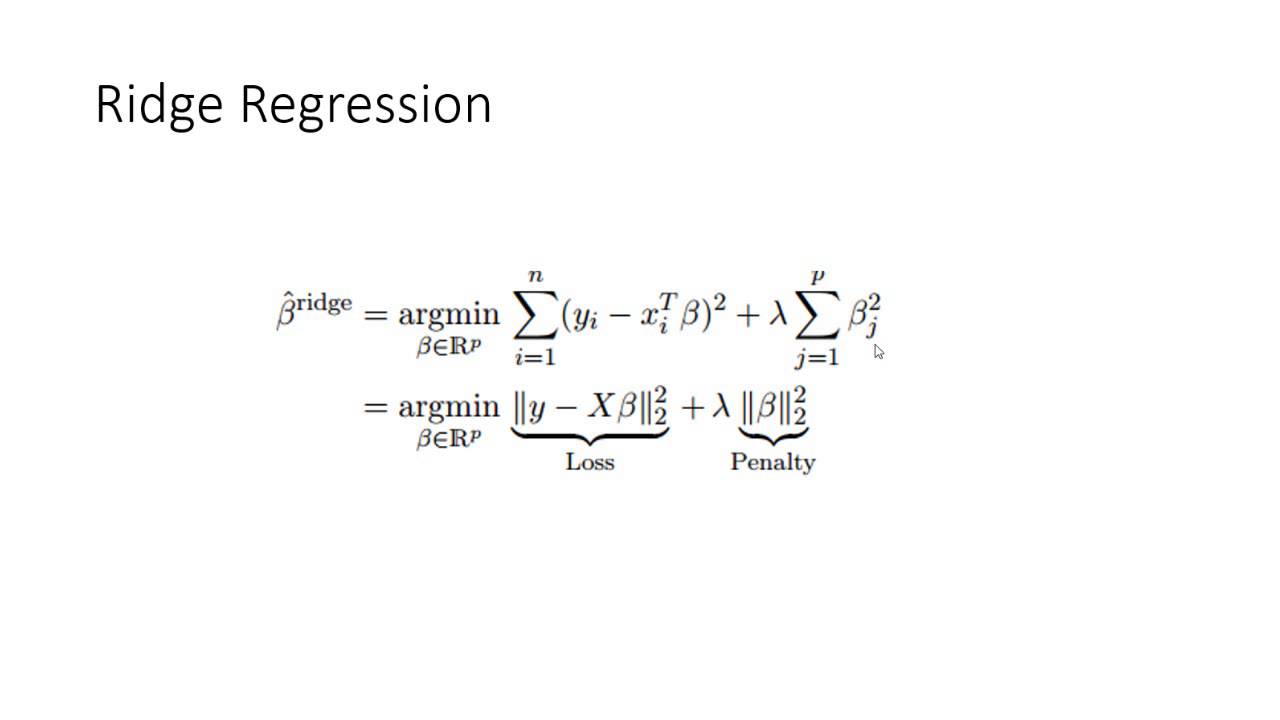
For example, in a simple regression problem (a single x and a single y), the form of the model would be:  
 y = B0 + B1\*x



.

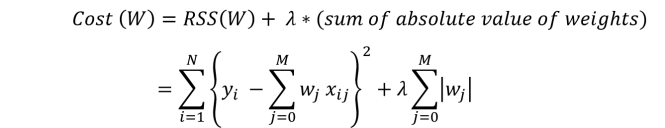
1. **Ridge Regression**

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It is hoped that the net effect will be to give estimates that are more reliable.



1. **Lasso Regression:**

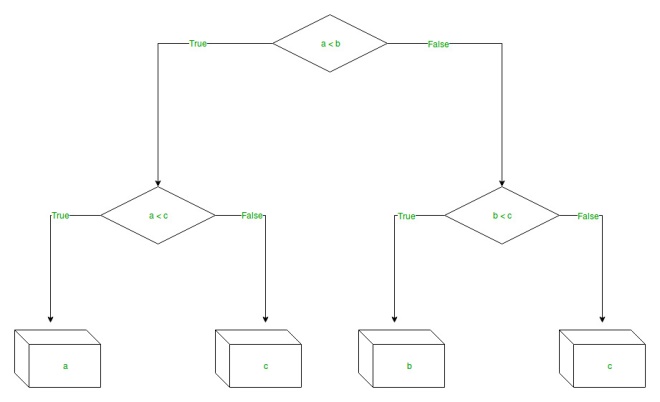
**Lasso regression** is a type of [linear regression](https://www.statisticshowto.com/probability-and-statistics/regression-analysis/find-a-linear-regression-equation/)that uses [shrinkage](https://www.statisticshowto.com/shrinkage-estimator/). Shrinkage is where data values are shrunk towards a central point, like the [mean](https://www.statisticshowto.com/mean/). The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of [muticollinearity](https://www.statisticshowto.com/multicollinearity/) or when you want to automate certain parts of model selection, like variable selection/parameter elimination.



1. **Decision Tree Regressor:**

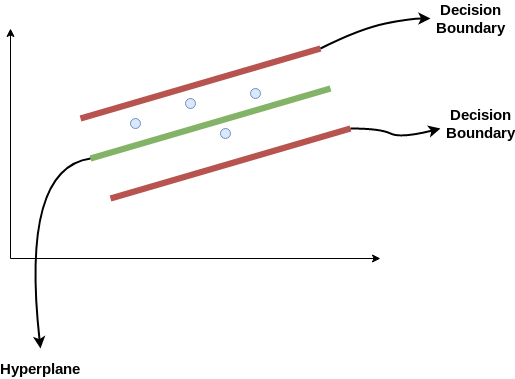
Decision tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs and utility.

Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables.



1. **Support Vector Regression:**

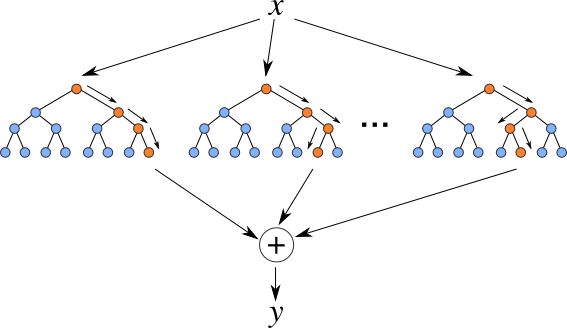
Support Vector Regression (SVR) uses the same principle as SVM The problem of regression is to find a function that approximates mapping from an input domain to real numbers on the basis of a training sample.



* **Kernel:** A kernel helps us find a hyperplane in the higher dimensional space without increasing the computational cost. Usually, the computational cost will increase if the dimension of the data increases. This increase in dimension is required when we are unable to find a separating hyperplane in a given dimension and are required to move in a higher dimension:
* **Hyperplane:**This is basically a separating line between two data classes in SVM. But in Support Vector Regression, this is the line that will be used to predict the continuous output
* **Decision Boundary**: A decision boundary can be thought of as a demarcation line (for simplification) on one side of which lie positive examples and on the other side lie the negative examples.

1. **Random Forest Regression:**

**Random Forest Regression is** a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, over fitting may easily occur, we must choose the number of trees to include in the model.



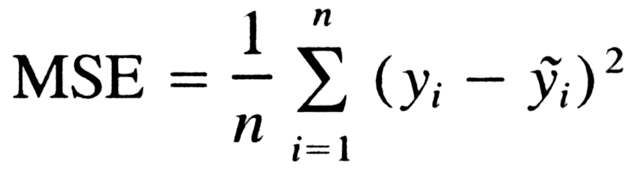
**6.2. Model performance:**

Models are evaluated on mean squared error

1. **Mean Squared Error:**

In [statistics](https://en.wikipedia.org/wiki/Statistics), the mean squared error (MSE)[[1]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-:0-1)[[2]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-:1-2) or mean squared deviation (MSD) of an [estimator](https://en.wikipedia.org/wiki/Estimator) (of a procedure for estimating an unobserved quantity) measures the [average](https://en.wikipedia.org/wiki/Expected_value) of the squares of the [errors](https://en.wikipedia.org/wiki/Error_(statistics)) that is, the average squared difference between the estimated values and the actual value. MSE is a [risk function](https://en.wikipedia.org/wiki/Risk_function), corresponding to the [expected value](https://en.wikipedia.org/wiki/Expected_value) of the squared error loss.

The MSE is a measure of the quality of an estimator. As it is derived from the square of [Euclidean distance](https://en.wikipedia.org/wiki/Euclidean_distance), it is always a positive value with the error decreasing as the error approaches zero.



**6.3. Hyper parameter tuning:**

Hyperparameters are sets of information that are used to control the way of learning an algorithm. Their definitions impact parameters of the models, seen as a way of learning, change from the new hyperparameters. This set of values affects performance, stability and interpretation of a model. Each algorithm requires a specific hyperparameters grid that can be adjusted according to the business problem. Hyperparameters alter the way a model learns to trigger this training algorithm after parameters to generate outputs.

I used Grid Search CV, for hyperparameter tuning. This also results in cross validation and in my case I divided the dataset into different folds.

1. **Grid Search CV-**Grid Search combines a selection of hyperparameters established by the scientist and runs through all of them to evaluate the model’s performance. Its advantage is that it is a simple technique that will go through all the programmed combinations. The biggest disadvantage is that it traverses a specific region of the parameter space and cannot understand which movement or which region of the space is important to optimize the model.

**7. Conclusion:**

That's it! I reached the end of my exercise.

Starting with loading the data so far I have done EDA , null values treatment, feature selection and then model building.

In all of these models the test error which was measured as Mean Squared Error was around 0.49 to 0.93. The model which gives less error rate is the best model.

After Hyperparameter tuning Random Forest Regressor provided with the less error rate during testing but during training it can be seen that the model was completely over fitting. Hence SVR is chosen as the best Regression model for predicting energy.

SVR performs quite well as compared with other models which provides an Mean Squared error rate of 0.368 During Training and While testing Mean Squared error rate is 0.57 which is best as compared with other models.