Part1

THE LIGHTING AND APPLIANCES USED 1.23 TIMES MORE ENERGY THAN DOMESTIC HOT WATER IN SPAIN, 1.28 MORE IN EUROPE and 1.74 MORE THAN IN UNITED STATES ACCORDING TO DATA FOR RESIDENTIAL SECTOR IN 2003

1) Focuses on energy use of appliances in a low energy house.  
2) Features on Data Removal to remove non-predictive parameters and feature ranking.  
3) 4 models were used to evaluate the results obtained.  
4) Data from kitchen, laundry and living room were ranked with high significance during the prediction.  
5) Those models which used only weather data as a parameter found that atmospheric pressure was the most relevant feature to be considered during prediction.

1) Electricity consumption depends on 2 main factors which are:

- Type and number of appliances.

- Use of appliances by the occupants.

2) Different outdoor and indoor parameters were considered while prediction.  
3) Different combinations of predictors are used in the 4 models to contrast and identify the significant ones in the data.  
4) During the process, research related to studies of appliances load in buildings and modelling are also included.

1) The purpose of this work is to understand the relationships between appliances energy consumption and different predictors.  
2) After the introduction, the paper continues with a description of the house and follows with a description of the energy meters and wireless sensor network.  
3) Then, series of Exploratory Data Analysis, Data Filtering and the training of models with different data subsets is shown and the evaluation of trained models in the test sets is presented.

1) Is the weather data from a nearby airport representative enough to predict the energy consumption through appliances?  
2) Can the temperature and humidity measurements from the wireless network help in energy predictions?  
3) From the data used in all the models which features are most useful in predictions?  
4) What effect does sub-metered energy measurement has on the prediction?

1) 75% of the data is used for training the models and the rest in testing them.  
2) Pair plots are created which shows the relationships between all the variables with the energy consumption of appliances in the training set. Psych package was used for the same.  
3) The highest correlation is observed between NSM and appliances with a value of 0.22. Also there is a positive correlation with appliances consumption and wind speed with a value of 0.09, higher wind speeds correlate with higher energy consumption by the appliances.  
4) Negative correlation was observed between appliances and RHout with a value of -0.15.  
5) It is seen that the energy consumption starts to rise around 6a.m and then around noon there are load surges. It is clearly observed that there is increase in demand for energy around 6p.m

1) Boruta package was used to carry out feature selection. Boruta package compares importance of features with importance of shadow attributes that are created by shuffling original ones.  
2) To test how many variables are optimal to minimize the RMSE, RFE(Recursive Feature Elimination) is used to select optimal inputs.

1) Initially, Multiple Linear Regression Model was trained which uses all the predictors and finds the appropriate slope quantifying the effect of each predictor and the response. The residuals were computed as the difference between the real values and the predicted values. The linear model could not identify the relationship between variables and energy consumption efficiently due to residuals being not distributed normally along horizontal axis.  
2) Support Vector Machines with radial basis kernel function was trained next which has numerical advantages which are helpful in this study.   
3) Random Forest which is a tree based model requires finding the optimum number of trees and the number of randomly selected predictors. But, RMSE was not changed after 300 number of trees and 18 predictors.  
4) GBM model improved the prediction by using first trees information, finding optimal parameters for number of trees and maximum depth of tree.

1) After training the models, each model has 30 results from 10-fold cross validation (CV) sets and 3 repeats.  
2) It was observed that RF and GBM models have very similar performance based on their RMSE and R-squared values and confidence intervals. RF and GBM models have very similar performance based on their RMSE and R-squared values and confidence intervals.

1) After concluding that GBM performed well and provided best RMSE and R^2 values this model was used to further study the performance of the model with different predictors, with a mix and match of different features and removing the irrelevant one.  
2) The GBM Model without lights predictor is accurate in comparison with the original GBM model yields R^2 value of 0.58 in testing set.

1) The consumption of appliances represents the highest percentage of electrical consumption with almost 75% of monthly consumption.  
2) During Exploratory Data Analysis, the correlation between some of the features was high for ex: T5 and T8 and so only the most representative ones can be used for the predictions.  
3) The relation between Humidity and appliances consumption was found to be very good, as the number of occupants in the room increases, the humidity in the room increases and hence energy consumption increases.  
4) The data filtering is important as it diminishes some of features which have no use in predictions. According to RFE, the six parameters can significantly reduce the RMSE which are: NSM, lights, pressure, RH5, T3 and RH3.  
5) When looking at the ranking for the GBM model with no light information (see Fig. 17), it can be seen that the top predictors are the NSM, Pressure, RH1, RH2, RH3, RH5, T6, RH6, RH4, RH9, T8, T4, and T2 which means that information from the kitchen, living room, laundry room, bathroom, outdoors, office, and bedrooms are the most important.

1) As the analysis was performed on a single house, important information and features could be identified analyzing different houses which different factors such as no of occupants, no of pets, location.  
2) The models would have performed better considering the factors like if the weather station was near to house, number of sensors and location of sensors.

1) The statistical data analysis has shown thought-provoking results in both the exploratory analysis and in prediction models.  
2) The weather data from the nearby weather station was shown to increase the prediction accuracy in the GBM models.  
3) The study has found curious relationships between variables. Future work could include considering weather data such as solar radiation and precipitation.  
4) Also, the occupants activity could be tracked using different sensors which would gain information about average time spent in a specific room.

Part2

Load management uses two types of load control:

* #### Direct Control : When the loads are cut off on command.
* #### Control by cost : Meeting the demand by varying the cost of Energy consumption. Just like applying Surge pricing during peak travel hours.

According to the survey of Energy consumption by sector in European Countries in 2007, The households and services plays the major part in Energy consumption (37.1%)

Types of Predictors :

* **The “will always consume” predictor:**

This predictor assumes that energy is consumed by permanently by the appliance

* **The "will never consume" predictor :**

This predictor assumes that energy will not be consumed by the appliance in the next day

* **The ARMA predictor :**

The Auto Regressive Moving Average(ARMA) predictor assumes the current value of a time variable as a function of its past values

**Proposed Predictor Model:**

The proposed model is generated on the basis of the assumption that variation of electriciry consumption varies with inhibitant's day to day behaviour due to various timely factors and thus can be seen as having a random probability distribution.

**Improving the prediction precision**

A time based study is done while mining the data that shows varition of electricity consumption based on seasons, time of day, weekdays or weekends etc.

For example:

There is a greater demand of energy during winters due to the use of heaters, A halogen bulb will consume electricity during the evening hours etc.

To increase the weightage of time in prediction, the data is segmented based on the days where energy consumption is similar. A K-Means clustering algorithm is applied for this purpose.

Two clusters are obtained based on the result of iterative K-means clustering, namely, C1 and C2 which shows weekday's and weekend's data respectively. The original data is then divided based on obtained clusters C1 and C2 and predictor is applied to get the prediction precision

Fig 13-14 show that using segmentation, the prediction precision is improved significantly.

**4. Global study of the services in the house:**

When the proposed predictor is applied on different appliances used in house for a 1 year period, The prediction precision obtained is expectedly different for different appliance.

For example, prediction precision for Refrigerator or Freezer is lower than other predictor models for a range more than 14 days, Which is as expected because of changing season requirements. Whcih implies using a shorter range for prediction using proposed predictor model. The same is show in Fig 15 and Fig 17 below

Another example of Television is show in Fig 18, for which the proposed prediction precision is higher, as the usage and energy consumption by this appliance is not based on season

Part3

The world wide energy consumption by Buildings is estimated around 30 % of total energry consumption. Any efforts in reducing energy use or increase energy efficiency for this sector will reduce reliance on global energy remarkably.

There are 3 widely used methods to predict building energy usage:

* #### Engineering Method: This method uses Thermo dynamic equations to measure physical behaviour and there interaction with the environment condition to estimate energy use. This method is also know as White Box method, as details of every component in the building is known.
* #### AI Based Method: This method predicts energy use without knowng internal relationship of the buildinng and its individual components, also refered to as Black Box Method.
* #### Hybrid Method: This is also known as grey box method as it uses both white box and black box methods.

Both Engineering and Hybrid method require detailed data of building components which makes it difficult as most of the existing buildings donot have compomnents from which this data can be extracted. This makes AI based method important as it dosent require data of all building components

Most of the previous AI based research models used data from educational and research, and commercial building due to data availability and, potentially, easier access to the available data

AI based prediction models consist of 4 main steps:

Data Collection -> Data Pre-processing -> Model Training -> Model Testing

The evaluation of models are done using performance indicatiors like RMSE, MAPE, Coefficient of Determination(R2)

91 % of research work of AI based model is based on single prediction models like:

* Multiple Linear Regression(MLR) : Advantage of this model is that No tuning of features is required, Ease of use compared to other models Limitations include poor accuracy for non linear models
* Artificial Neural Network(ANN) : Advantages of using ANN is its ability to detect complex non linear relationships Limitations include poor accuracy when changes are made to the building components affecting energy use
* Support Vector Regression(SVR) : Advantages include attaining high level of prediction accuracy once appropriate parameters are selected for model training Limitations are that there is no standard way of determining kernel function/ parameter selection

Advantages of Single Prediction Model are :

* Reliability
* Ease of implementation
* Fast computation

The Ensemble model was introduce in 1990's, but research on using this model for Energy usage predictions of buildings is done only after 2014. It uses multiple learning algorithms to obtain better prediction performance. in this model the data collection and data pre-processing is similar to single prediction models, but the difference lies in model training and testing.

Ensemble models can be categorized as follows:

* Homogenous Models : Here the learning algorithm or the base learner is kept same, but different samples are taken for training the model.
* Hetrogenous Models : Here the training set is kept constant and different base learners are provided with this training set.

Ensemble model follow the below steps:

Input feature identification -> Data Monitoring and Pre-processing -> Learning Algorithm selection -> Base model generation -> Model Integration

Advantages of Ensemble models:

* Highly reliable models
* Higher prediction accuracy is obtained of right set of learning algorithms are selected for base learners

Drawbacks of Ensemble Models:

* Requires more calculation time
* Requires high amount of knowledge of different models
* Prediction accuracy highly depends on base model selection