**Power Consumption Prediction**

**PROJECT REPORT**

**Submitted by**

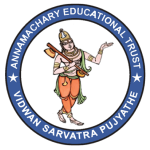
**The Shield Crew**

* **L.Srinivasreddy**
* **K.Naveen**
* **D.Rakesh**
* **V.Aryan Akshay**
* **B.Sai Manoj**

***In partial fulfilment for the award of the Certificate***

**Of**

**SUMMER INTERNSHIP PROGRAM**

**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village, Renigunta Mandal, Tirupati, Andhra Pradesh 517520**

**July 2019.**

### BONAFIDE CERTIFICATE

This is to certify that the project entitled ”**PROJECT TITLE**” submitted by **Team Members Names** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### Signature of Supervisor                                       Signature of Head of the Department

 Akshay Kumar Kothuri, M.Tech       Mrs.B.Rupa Devi M.Tech, Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS, Tirupati.

Power Consumption Prediction

Abstract:

Total Power Consumption is available for most households. However, the electricity consumption related to different end uses, e.g. space heating, lighting and services from household appliances are usually not metered. Metering data are costly to achieve, and in this paper we study two methods for end-use estimation, which can be applied on household data for appliance holdings, demographic and economic variables. The first method is the engineering model, which has been used to calculate the so far only documented Norwegian end-use results applied on data from a Norwegian energy survey. The second method is an econometric conditional demand model applied on data from the same survey. We compare the numerical results from the two models and give some recommendations regarding choice of end-use approach and what questions to implement in household surveys designed to disaggregate electricity consumption.

**Introduction :**

* **The Power Consumption’ is all about total power consumed by a particular house hold or total power consumed globally over span of time . Power consumption can help people to estimate how much power they are utilizing to their household . Power consumption can be calculated using following criteria like : No of hours usage , No of days usage , Capacity of appliances[watts] in Kwh**

**PYTHON:**

Python, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks. The simple syntax rules of the programming language further makes it easier for you to keep the code base readable and application maintainable. Main reasons to use python language is:

1. Readable and Maintainable Code

2. Multiple Programming Paradigms

3. Compatible with Major Platforms and Systems

4.  Robust Standard Library

5. Many Open Source Frameworks and Tools

6. Simplify Complex Software Development

7. Adopt Test Driven Development.

**Machine learning:**

Machine learning is an subset of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Basically Machine Learning is of three types:

1.Supervised Learning: Learning from characterized data.

2.Unsupervised Learning: Learning from raw data.

3.Reinforcement Learning: Learning from self mistakes (or) self learning data.

Predictive modelling is the way of building a model that is capable of making predictions. The process includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions. Predictive modelling can be divided further into two areas: Regression and pattern classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables. In contrast to regression models, the task of pattern classification is to assign discrete class labels to particular data value as output of a prediction Here we used multiple linear regression algorithm to predict the future power consumption rate.

**Steps involved in Machine learning:-**

Steps involved in the machine learning is shown in the below figure

1. Data Collection

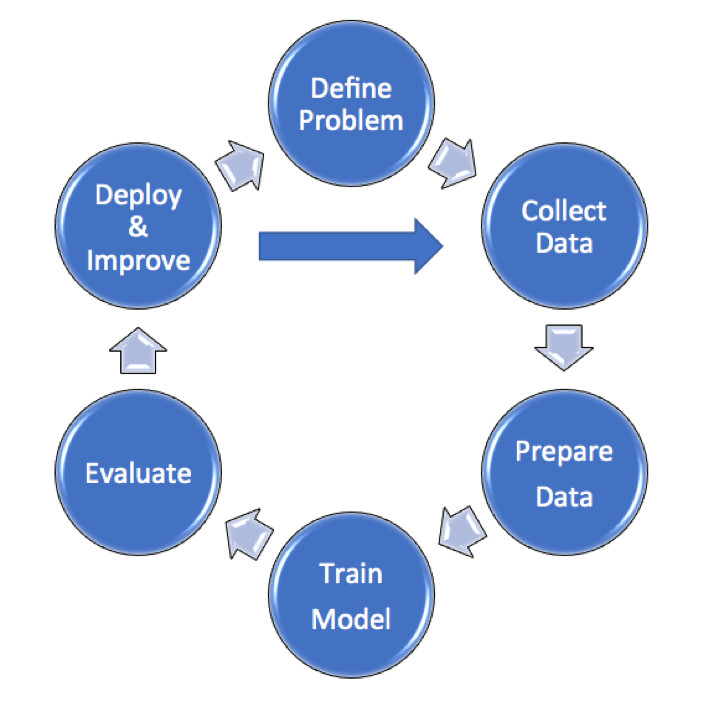
2. Data Wrangling

3. Analyse the data

4. Train the algorithm

5. Test Algorithm

6. Deployment



**Data collection:**

Here we collect the data from the power departments and we use previous data and predict the future power consumption

**Data Pre-processing:**

In this phase, the data is prepared for the analysis purpose which contains relevant information. Pre-processing and cleaning of data are one of the most important tasks that must be one before dataset can be used for machine learning. The real-world data is noisy, incomplete and inconsistent. So, it is required to be cleaned.

**Extraction of Feature Set/Training Data**

The feature sets and training set that has obtained by using any of the methods will be used for the implementation of machine learning algorithms.

**Implementation of Machine Learning Algorithm on Feature Set/Training Data:**

**Regression:**

A Regression model is created when we want to find out a number – for example: How many days before a patient discharged from hospital with a chronic condition such as diabetes will return.

**Classification:**

To determine a label or category – it is either one thing or another. We train the model using a set of labelled data.

**Testing of Data:**

Testing of data is done based on training model which is classified using supervised learning algorithm. Evaluation of the total responses for every question and determine the polarity of feedback received in context of the given data.

**Multiple Linear Regression:**

Multiple Linear Regression is the most common form of linear regression analysis.  As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables.  The independent variables can be continuous or categorical (dummy coded as appropriate). The multiple linear regression algorithms is having one and more independent variables and one dependent variable. Here it consist of hyper plane i.e. is all the straight lines joined together is called hyper plane. The iterations are performed up to a0=0. The equation of the multiple linear regression y=a0+a1\*x1+a2\*x2........

**1.2. Objective of research**

The above problem made me to go for a research about how can we predict the power consumption for making easier. Through many documentation and cases, it came out that machine learning and data science can make the work easier and faster. The objective of this project is to analyze dataset which consist of power criteria per span of time and predicting the power consumption which may happen in future depending upon dataset consisting of previous data.

**1.3. Problem Statement**

Finding out the future prediction of Power Consumption by using Machine learning algorithm using python as core.

**2. Review of literature**

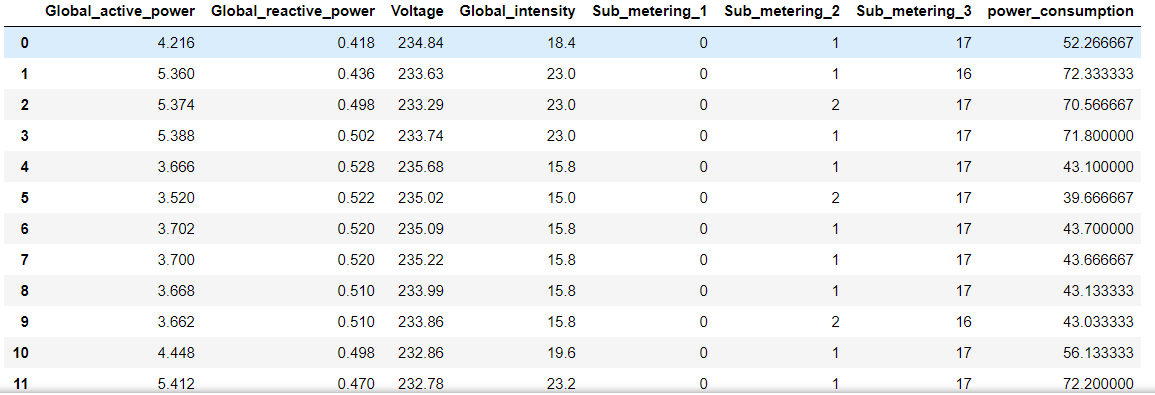
[**www.machinelearningmastery.com**](www.machinelearningmastery.com)

The dataset used in this project is taken from machinelearningmastery. The dataset obtained from machinelearningmastery is maintained and updated by the power consumption department. Power consumption dataset from machinelearningmastery is used in CSV format.

**3. Data Collection:**

The data is collected and uploaded in the Jupiter notebook in Anaconda navigator. we use pandas , numpy and we declare pandas as pd and numpy as np. And we assign dataset to dataset variable by using "project1.read\_csv" following with dataset name and we print the dataset. And next we slice the data with independent variables as x and dependent variable as y. here the independent variables are Global\_active\_power,Global\_reactive\_power,Voltage,Global\_Intensity,Sub\_metering1,Sub\_metering2 and Sub\_metering3.

and Power Consumed as dependent variable. Here we use "iloc" to slice and we convert the data set into arrays using ".values".



**4.1 Statistical techniques:**

. . Multiple linear regression is the most common form of linear regression analysis.  As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables.  The independent variables can be continuous or categorical (dummy coded as appropriate) The multiple linear regression algorithm is having one and more independent variables and one dependent variable. Here it consist of hyper plane i.e. is all the straight lines joined together is called hyperplane. The iterations are performed up to a0=0. The equation of the multiple linear regression y=a0+a1\*x1+a2\*x2.........

From "sklearn.model\_selection" we import "train\_test\_split " ,and we declare variables x\_train, x\_test, y\_train,y\_test and these are assigned to train\_test\_split of x,y with test\_size =0.2, and we declare here random\_state as zero.

Random Sampling (Train and Test)

Training Sample: Model will be developed on this sample. 70% or 80% of the data goes here.   Test Sample: Model performances will be validated on this sample. 30% or 20% of the data goes here.

**4.2 Data modelling:**

From sklearn.linear\_model we import the linear regression and we assign "lr" variable with linear regression and then we give the splitted values to the model with "fit" .to predict the values we use predict. Here our input independent attributes are:

Global\_active\_power,Global\_reactive\_power,Voltage,Global\_Intensity,Sub\_metering1,Sub\_metering2 and Sub\_metering3

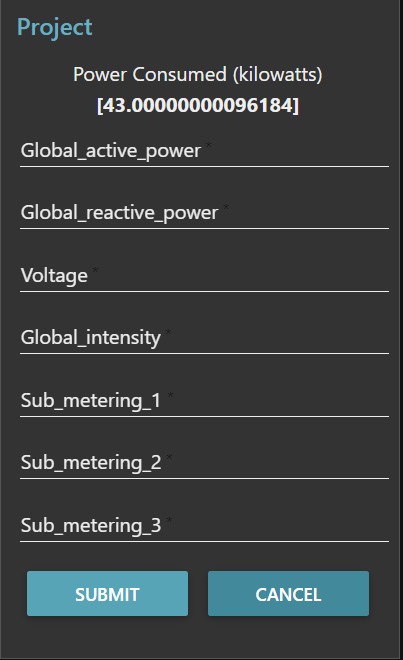
We get Power Consumption as Output.

4.3 PROJECT LINK:

<https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/2eef6635-e92a-4d17-8b23-cf954b5937c3/view?access_token=0e603f9007d8f47958725b52bee7c60f4a850c6d8dafeb27d18e298225997fb3>

4.4 Prediction Form:

<www.tinyurl.com/powerconsumption>



**5. Conclusion:**

With the help of machine learning technology, it has become easy to find out relation and patterns among various data’s. The work in this project mainly revolves around predicting the future power consumption. Using the concept of machine learning we have built a model. We generated the future power consumption value which helps the power departments and power grid corporations and electrical departments to explore the power consumption prediction.

* The household power consumption dataset that describes electricity usage for a single house over a period span.
* How to explore and understand the dataset using a suite of line plots for the series data and histogram for the data distributions.
* How to use the new understanding of the problem to consider different framings of the prediction problem, ways the data may be prepared and modeling methods that may be used.