**Abhishek Srivastava**

**abhisheksrivastavabbn@gmail.com**

1. Coding ⇒ Exploratory Analysis ⇒ Hourly Data ⇒ Timestamp and the load in MW
2. Final ⇒ Timestamp data (day wise) and weather data ⇒ Excel Sheet is available

**Dataset (Saturday 27-11-2021)**

1 Hourly Prediction (Load Data) ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/ESLn6YfaQOVNtTTgBv3J71oBk1pEt9j1qmFGzKAj9wjQZQ?e=A1jKXQ))

2 Day Wise prediction (Load Data) ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/Ed3FMQiXqTlAsE0MfsHjuOUB01JXkwPuvwQ1yIjbY6rjTA?e=ALApkZ))

3 Hourly Prediction (Load Data) + weather data ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EQb4Ah0a29FLvNq1onZqsg0B-RnKhKZdMMo76SqPpCJpXg?e=4qNu9i))

4 Day Wise prediction (Load Data) + weather data ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EYY5ehUu9cJNpte9CbWnVnEBvhsMInHMDXOJIAGacNDg4A?e=uSa1WZ))

5 Hourly Prediction (Load Data) + Holiday or Not ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EQb4kaOZXwJAge4RXXxe77cBKZFVuwnpcOZEViWXQcKu_Q?e=Xkjqro))

6 Day Wise prediction (Load Data) + Holiday or Not ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EZIqTL5-AdZNr1b5mKggj6wB2sSbj5ujYnlfU-s--VTDxw?e=B9oIUC))

7 Hourly Prediction (Load Data) + weather data + Holiday or Not ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EdZw-MsLIZhPn1TiJ1j0NIABNDXsOfHsRZDCppw_IBp9Tw?e=DvxJCW))

8 Day Wise prediction (Load Data) + weather data + Holiday or Not ([Link](https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EQWuWe58FAFHka6_m8h5rKgBBrk3ADLtjPMjdSa9IhVJvA?e=n4l2D9))

* Dependent Variable MW
* Rest all would be Independent Variable

Here is the [**Notebook**](https://colab.research.google.com/drive/1pGPAcjHg_KkYKXVVtAjxyImfYIFG-CAy?usp=sharing) **used for Dataset Manipulation**

**Models -** [**Bokeh Visualization Notebook**](https://colab.research.google.com/drive/10G8We5HrTElQAEK7_ZolQShazDRF4zzL?usp=sharing)

Histogram

Year - Month - MW

Trend Analysis

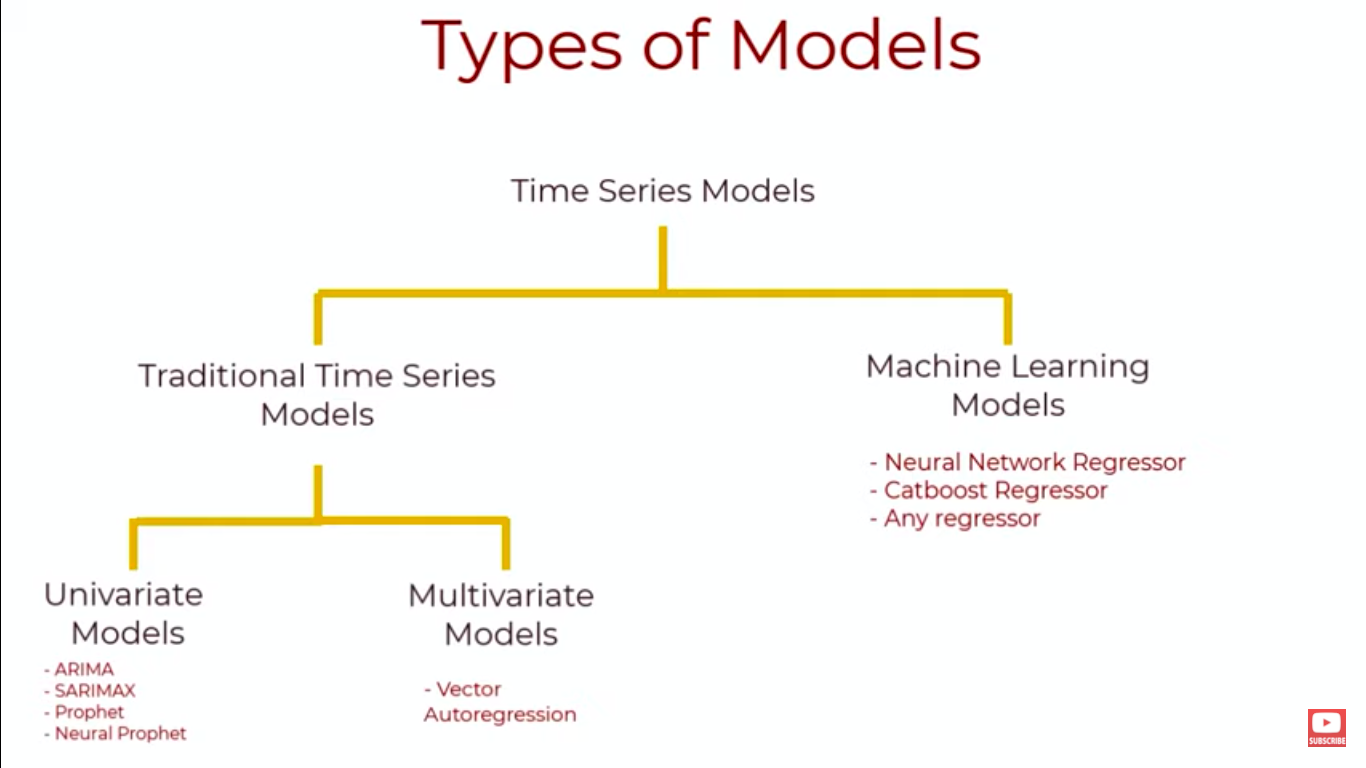
* Plot MW Vs holiday (Weekend)
* MW Vs Working day
* MW Vs Festival
* Exploratory and Interpretation

Bokeh ⇒ Visualization package

* 2020 removal
* Study models (ARMA, ARIMA, etc)
* Inference

**Notebooks:**

* AR - [Link](https://colab.research.google.com/drive/1WkmVWwr_W0ZNk5BgyS8lMZKhEIcFyM_t?usp=sharing)
* ARIMA - [Link](https://colab.research.google.com/drive/1c0Ag4dAGLpqbZvraAZrItvr01T3yFhGl?usp=sharing)
* Prophet - [Link](https://colab.research.google.com/drive/1I1ysu1fWaYErwFkSHKPkizkiJPF1jxB9?usp=sharing)
* Linear Regression - [Link](https://colab.research.google.com/drive/1g9eFjdB-jP65EHNcAZDMjegQAMfT3tfI?usp=sharing)
* KNeighborsRegressor (KNN) - [Link](https://colab.research.google.com/drive/1RBZea1XFxBa1nmWM4ObnXRUuCyIqFRsZ?usp=sharing)
* RandomForestRegressor - [Link](https://colab.research.google.com/drive/1RIQhGSyP-1QmXwNc7ImdXoo4oDByy_g-#scrollTo=DzuWGFyYCOxv)
* Xgboost - [Link](https://drive.google.com/file/d/1L0zR62P0qAs0Ixy8-lWz1yLDjBZE-4uK/view?usp=sharing)

**Approaches:**

* Similar Day lookup
* Regression-based approach
* Time series analysis
* Artificial Neural Network(ANN)
* Experts System
* Fuzzy logic
* Support Vector Machine
* Delphi Method
* PCA
* Deep neural network

**Others:**

* Holt's Two parameter method
* Brown One Parameter Quadratic method
* Chow’s Adaptive method
* MAPE method (An efficient approach for short-term load forecasting using historical data)
* Prediction by chaotic dynamic reconstruction using the Grassberger-
* Procaccia algorithm and the least squares regression method are applied to obtain the value of the correlation dimension that will be the basis of the FNS model.
* Semi-parametric regression method
* Neuro-fuzzy structure that can be defined as an ANN
* Multi-Layer Perceptron (MLP)
* Radial Basis Function (RBF)
* Reptree
* Gaussian process (GP)
* Gradient Boosting Regression Trees (GBRT)
* Classification and Regression Tree (CART) analysis
* Feedforward neural networks (FFNNs)
* LSTM networks.

**Statistical analysis**

AR

MA

ARMA

ARIMA

**Machine learning**

Linear Regression

**Deep learning**

LSTM

RNN

**Holiday Dataset**

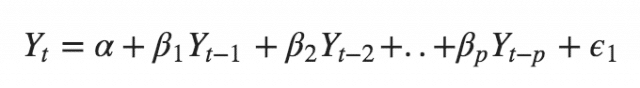
**Notebook:** <https://colab.research.google.com/drive/1lLnSuZd_goQiJKtWXOukViHeIW5nhqdR?usp=sharing>

**Resultant CSV File:** <https://vitbhopalacin-my.sharepoint.com/:x:/g/personal/abhishek_srivastava2019_vitbhopal_ac_in/EQvjub1Rlf9Ig-lyw4lrI6EB76TjQ1OREhGxv5SvM4fwLg?e=MNf2rS>

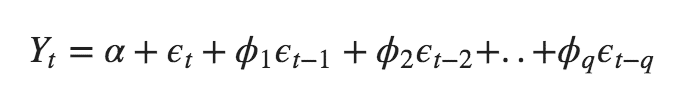
| **ARIMA** | | | |
| --- | --- | --- | --- |
| DATASET | P | D | Q |
| Hour vs MW | 5 | 1 | 1 |
| Day vs MW | 1 | 0 | 3 |

**Points to note down:**

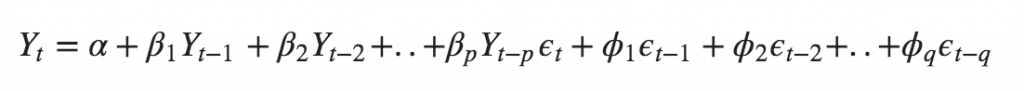
* Implemented AR and Linear Regression Model.
* There are three types of pattern in time series data:
  + Trend only
  + Seasonal only
  + Trend and Seasonal
* Our dataset follows a seasonal pattern
* A pure **Auto Regressive** (AR only) model is one where Yt depends only on its own lags. That is, Yt is a function of the ‘lags of Yt’.



* A pure **Moving Average** (MA only) model is one where Yt depends only on the lagged forecast errors.



* An **ARIMA** model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So the equation becomes:



* ARIMA has three components:
* p: Number of autoregressive lags,
* d: Order of differencing required to make the series stationary,
* q: Number of moving average lags.
* s: Measure of seasonality
* We perform the Augmented Dickey Fuller test (ADfuller()) to know the stationality of data. It returns a value p. Lower value of p means better results.
* Pacf and acf tells how correlated different time period are
* pacf => Partial Correlation - Direct effect of a perticular time period to current time period
* acf => Partial Correlation - Indirect effect of a perticular time period to current time period

Requirements:

* Stationary
  + Constant mean
  + Constant variance
  + No seasonality (pattern)
* How to verify stationality?
  + Visual inspection
  + Global and local check
  + **ADFT**
    - **ACF**
    - **PACF**
  + KPPS Test
* ADFT TEST
  + There are two Hypotheses: H0 and H1
  + P < 0.05 => Rejecting the null hypothesis (h0) and the data is stationary
  + P = 0.39089034821
    - To rectify the issue, we consider differencing
* Example: AR
  + Yt = c1yt-1 + c2
* **AR** 
  + ACF => **to obtain q value** : both direct and indirect effect of previous time lags
  + PACF=> **to obtain p value** : both direct effect of previous time lags
* **AR +** I + **MA** Model
* P + d + q

ARIMA =. AR + I + MA

**ARIMA (p, d, q)**

**P => AR (PACF)**

**D => order of differentiation**

**Q => MA (ACF)**

* With ARIMA, we have inbuilt algorithms to obtain the optimised value of p, d, q
* **In ARIMA, we can determine the value of (P, d, and q) using the auto\_arima function**

## Seasonal ARIMA(SARIMAX)

* ARIMA on Seasonal Dataset