Forecasting Household Power Consumption

1	Problem Statement

Forecasting household power consumption

2	Explanation

This is a time series problem where we need to forecast power consumption based on past consumption of power in household.

3	Literature Review

In time-series there are many algorithms used to forecast. From a simple moving average to more complex model like ARIMA. Being ARIMA most used and accurate algorithm we are using ARIMA in our model. When we say forecast, then this indicates toward a regression problem too. Some times only applying good model like ARIMA doesn't full fill all scenario present and this motivates us to use some more complex model using ARIMA with regression technique.

4	Data

date: Date in format dd/mm/yyyy
time: time in format hh:mm:ss
global_active_power: household global minute-averaged active power (in kilowatt)
global_reactive_power: household global minute-averaged reactive power (in kilowatt)
voltage: minute-averaged voltage (in volt)
global_intensity: household global minute-averaged current intensity (in ampere)
sub_metering_1: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).
sub_metering_2: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.
sub_metering_3: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

5	Deliverable

To make a model which will forecast power consumption in household.

6	Evaluation

Root Mean Squared metric is used.

7	Data Ingestion

We are reading data using pandas library as pandas data frame.

8	Data Analysis

Data contains missing value.
Outlier treatment is required.

9	Data Munging

Some rows of data was ignored as it contains string value while other are all numeric.

10	Data Exploration

Missing value is imputed using mean method as all feature are continuous.
Due to presence of outlier data was treated with mean and standard deviation keeping normal distribution hypothesis.

11	Feature Engineering

Active power was used in model1.

12	Modeling

Two models are used and compared.
1. Simple model using only ARIMA and active power.
2. Combination of ARIMA and linear regression is used taking all feature in consideration.
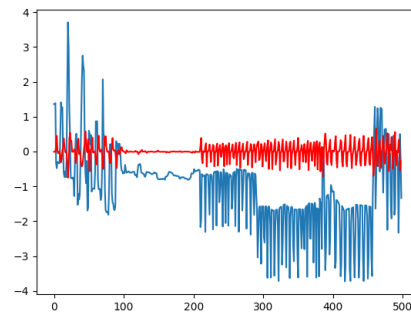
13	Optimization

Manual optimization was done for ARIMA hyperparameters.

14	Prediction

RMSE is 1.2 for model 1.

15 | Visual Analysis

Here red line is showing prediction.

16 | Results

From above we can see that initially model predicted accurately but later it while maintaining pattern its displaced from actual value.