**Name:** Burla Ramu

**Email address:** ramuburla7250@gmail.com

**Contact number:** 8714350696

**Anydesk address:**

**Years of Work Experience:** **2 years**

**Date:** **21st Jul 2020**

**Self Case Study -1:** M5 Forecasting - Uncertainty(Estimate the uncertainty distribution of Walmart unit sales)

---

## Overview

1. To estimate the uncertainty distribution of unit sales of various products sold by Walmart in the USA as precisely as possible.
2. The Dataset given has train data for 1913 days and next 28 days for validation and the next 28 days for evaluation.It shows the number of units sold per day. The prices data shows the product price every week.
3. We need to find the prediction intervals for medians and quartile of 50,67,95 and 99% according to the competitions guide

---

## Research-Papers/Solutions/Architectures/Kernels

1. https://www.kaggle.com/c/m5-forecasting-uncertainty

Got the overview and also this competition is hosted by the M Competitions.

2. https://machinelearningmastery.com/how-to-develop-a-skilful-time-series-forecasting-model/

   As, I am new to time series, this actually gave me lot of information on how to deal time series and which all machine learning algorithms I can use and which algorithms are good for time series and this is an eye opening on how to perform machine learning on time-series data.

3. https://www.kaggle.com/allunia/m5-sales-uncertainty-prediction

   There is a new machine learning technique that I found here called Prophet from facebook where we can do time series forecasting in this.

4. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/#:~:text=ARIMA%2C%20short%20for%20'Auto%20Regressive,used%20to%20forecast%20future%20values.

This has a very clear idea about ARIMA model and how to make the given series into the pattern so that we can use the ARIMA model.If the given model is seasonal we can sue SARIMA. Also we can use auto_arima(), which actually takes all the possible order of AR(p), differencing(d) and order of MA(q) values.

---

**First Cut Approach.**

The given data has unit selling of products in 3 states - Texas, California and Wisconsins. Out of which three types of product categories are there- Food,

Household and Hobbies.Over all the data is from 10 stores from three different states as mentioned above.

1. Understand and explore the datasets given. Check all the columns and need to find out a sequence or pattern.
2. Based on the pattern, if it is a seasonal based time series or normal time series we use the necessary models.
3. We need to preprocess the data by combining the necessary datasets and need to work on one hot encoding for the categorical features.
4. Once the necessary preprocessing gets over based on the data, will train machine learning models and time series based models.
5. We can also tune hyper parameters based on the model we train to improve performance.
6. We can also try ensemble models based on stacking techniques to improve the model performance.

---