**Problem statement**:

Prediction of crop prices is one of the most important task to ensure efficient crop planning and food safety in the country.

i. The historical data for prices in district “Agra” of state “Uttar Pradesh” are reported daily on [Agmarknet](https://agmarknet.gov.in/).

ii. Prices for a particular date (say 20 Mar’2021) can be extracted from a URL on Agmarknet:

https://agmarknet.gov.in/SearchCmmMkt.aspx?Tx\_Commodity=24&Tx\_State=UP&Tx\_District=1&Tx\_Market=0&DateFrom=20-Mar-2021&DateTo=20-Mar-2021&Fr\_Date=20-Mar-2021&To\_Date=20-Mar-2021&Tx\_Trend=0&Tx\_CommodityHead=Potato&Tx\_StateHead=Uttar+Pradesh&Tx\_DistrictHead=Agra&Tx\_MarketHead=--Select--

1. Python script to fetch data: The python script for data extraction is in the “\*.ipynb file”
2. i) From the plots in .ipynb file it can be seen hat the sale of potato is raising from 3rd quarter of the year (nearly from October) and get to its peak price in December 2020 , and falls drastically during the end of the year or starting of new year.

ii) The markets like khairgarh,Samsad and jarjar has almost constanst raising in their prices during the 4th quarter of year , this could be due to the lack of supply of potatoes or the data for the price for those month are missing.

iii) The highest price observe was for the Archena marktet. The seller can get he most profit by selling their potatoes in Archena market, its the most expensive market among all the markets listed. And the cheapest marke available is for the jarar.

1. i) Data Preprocessing/cleaning:

1. Drop the columns like Grade which is same for all the data.

2. Get the percentage change in the min and max price of the commodity to observe the max fluctuation in price in different markets.

3. Get the average of differen price columns to observe what is the average price of the commodity over the year in differnt market and what changes occur due to change in price of one market to other.

4. Look at the dates or months in which prices are getting down , which can be due whethear condition or natural calmity.

5. I would look for the price fluctuation due to the “variety” of comodity for among different market.

6. If not specifically required for potato , i’ll also observe the price changes for other commodites , and compare is their specific pattern for the raising in price of potatoes or its same for other commodities, and find some more data for further conclusions.

7. I’ll look for other commodities to conclude why their is constan rate of increment for market like Khairagarh,Samsabad and Jarar , that whether their no supply for this particular commodity or the data is not recorded for those time period.

8. Find the correlation among the prices for different commodity to get best feature for the traning model.

ii)Features: If required i would create new features like precent changes in price and change in maximum and min prices. I would also convert commodity and variety to one hot enconding . I would find the correlation of feature column with target price column , to get best features to train model over . So the features would be market,vareity ,data, price change ,%change in price.

iii) We can predict the max cost of the commodity for different market for the particular month due to percentage change in prices and fluctuation , this could help the farmer or seller to sell crops to best market and at price such that their would not be a chance of inflation and growth would be equal. So the target variable would be the price of commodity on a date or % change in the price for the month.

iv) I could use he both multivariate/linear regression as per target variable, stochastic gradient descent,random forest(ensemble).

v) The target variable is somewhat ambiguous over the given set of data that we can have the price as the target variable and market,variety and date as the features , this might give result if their is any trend over the month or season but we need some features like profit or selling in order to predict prices in better results.