**A Framework for Time-series Data by Analysing the Quality of Time-series Data**

**[Crop - planning]**

Agriculture is the backbone of the economy of a country. When determining a crop price, farmers should get a decent price for their products and also customers should be able to purchase the products for an affordable price. If the weather and the demand of the crop for future is forecasted, the above goal can be achieved, can get a considerable profit for the crops and also able to avoid the shortage of crops. A stable and accurate ML/DL model helps to forecast the future demand and weather. In this project we propose a mode for robust crop price prediction by analysing historical market place data and weather data.

A time series is a collection of observations made sequentially in time. Time-series forecasting has been studied under this crop price prediction scenario. To forecast time-series specific features mainly two categories of model are proposed.

* **Univariate model** : Can only model the endogenous variables like MA(Moving Averege - q), ARMA(Auto Regressive Moving Average**),** ARIMA(Auto Regressive Integrated Moving Average**)**
* **Multivariate model :** Can model the exogenous variables along with endogenous variables like VAR(Vector Auto Regressive) model along with its variant such as elliptical.

Time-series forecasting also used under various scenarios such as stock price prediction, traffic forecasting, height of ocean tides, weather of a certain geolocation, etc.

First, data preprocessing was performed to deal with missing data and outliers(extreme values that are outside the range of what is expected and unlike the other data). There are several methods to fill missing values such as **Ffill(**taken from above rows in similar column and fill it in later), **Bfill(**taken from below rows in similar column and fill it in later), **Pad, Polynomial and Spline.** And also to remove outlier **Sorting method, Data visualization method, Statistical tests and Interquartile Range method.**

Secondly, the statistical analysis is performed to test whether the time-series data is non-stationary, strict-stationary, trend-stationary or differencestationary. The ADF and KPSS tests are used for statistical analysis. If the dataset is stationary, most of the time-series model can be implement for that data. If it is not stationary then the dataset have to convert to stationary.