i. The data pre-processing / cleaning techniques used are

- Changing the dates data type to datetime64[ns].

- Removing the serial number column and columns with single unique value.

- Replacing the 0 valued Min and max prices with the values of previous row value of the respective columns

- Label Encoding the categorical and datetime64[ns] data type valued columns

ii. The features used to create the model are Market Name, Variety,

Min Price (Rs. /Quintal), Max Price (Rs. /Quintal), Price Date.

iii. This data can be used to predict what the prices of these potatoes will be in the future as it is a continuous dependent variable from a number of independent variables in the data it can be classified as a regression problem, we can predict the min price, max price or the modal price using a regression model, since the modal price best represents the overall data, it was decided as a target variable.

iv. Algorithms tested for price prediction are Random Forest Regressor and XGBOOST Regressor out of these Random Forest Regressor is used for price prediction, since it is

- Easy to interpret

- Not sensitive to outliers.

- Works well on a large dataset.

v. The loss function that was used is Mean Squared Error and R-squared score is used as performance measure

vi. It was observed that,

- Both Min and Max prices are strongly positively correlated with Modal prices.

- Both Min and Max prices are in a linear relationship with the Modal prices.

- Most of the time the prices are around the value Rs.1000 /Quinta

- Other variety potatoes used to be sold more, but Desi variety became more popular over years.