REVA UNIVERSITY
Bengaluru, India

A Project Report on

"Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato"

Submitted in partial fulfilment for award of degree of

Master

of **Business Administration**

Submitted by

Suresha H.P

R19MBA10

Under the Guidance of

Mr. Krishna Kumar Tiwari

Machine Learning Architect, CoE AI/ML,JIO

REVA Academy for Corporate Excellence

REVA University

Rukmini Knowledge Park, Kattigenahalli,

Yelahanka, Bangalore – 560064

October, 2020

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, Suresha H.P hereby declare that I have completed the project work towards the first year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled "Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato" under the supervision of Mr. Krishna Kumar Tiwari, Machine Learning Architect, CoE AI/ML,JIO. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2019-2021.

Suresha H.P

Place: Bengaluru | Name of the Student:
Date: 24/Oct/2020 | Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to Certify that the Project work entitled "Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato" carried out by Suresha H.P with R19MBA10, is a bonafide student of REVA University, is submitting the first year project report in fulfilment for the award of Master of Business Administration during the academic year 2019-2021. The Project report has been tested for plagiarism, and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of PROJECT work prescribed for the said Degree.

Signature of the Guide
Name of the Guide
Guide

Signature of the Director
Name of the Director
Director

External Viva

Names of the Examiners

1. <Name> <Designation> <Signature>
2. <Name> <Designation> <Signature>

Place: Bengaluru

Date:

REVA UNIVERSITY
Bengaluru, India

Acknowledgement

I express my sincere thanks to Dr. Shinu Abhi and Dr. Jay Bharatheesh for their valuable guidance for my project. I sincerely acknowledge my mentor Mr. Krishna Kumar Tiwari, teachers, classmates, program office members, parents, family and friends who have directly and indirectly supported in my project.

I sincerely acknowledge the support provided by Hon'ble Chancellor, Dr. P Shayma Raju, Vice Chancellor, Dr. K. Mallikharjuna Babu, and Registrar, Dr. M. Dhanamjaya.

Place: Bengaluru
Date: 24/Oct/2020

Similarity Index Report

This is to certify that this project report titled "Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato" was scanned for similarity detection. Process and outcome are given below.

Software Used: Turnitin

Date of Report Generation: 24-Oct-2020 04:19PM (UTC+0530)

Similarity Index in %: 6%

Total word count: 6447

Name of the Guide: Mr. Krishna Kumar Tiwari

Swetha H.P

Place: Bengaluru

Date: 24/Oct/2020

Name of the Student:

Signature of Student

Verified by:

Signature

Dr. Shinu Abhi,

Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	AUTO ARIMA	“Autoregressive Integrated Moving Average”
2	VAR	“Vector autoregressive model”
3	XG Boost	“X Gradient boosting model”
4	RNN	“Recurrent Neural Network”
5	LSTM	“Long short-term memory”
6	SVR	“Support vector Regression”
7	AI	“Artificial Intelligence”
8	ML	“Machine learning”
9	MSE	“Mean Squared Error”
10	MAPE	“Mean absolute percentage error”
11	CRISP-DM	“Cross industry standard process for Data mining”
12	ACF	“Auto correlation function”
13	PACF	“Partial auto correlation function”
14	API	“Application programming interface”

List of Figures

No.	Name	Page No.
Figure 1.1	National agricultural commodity map	12
Figure 5.1	CRISP-DM lifecycle	16
Figure 5.2	RANDOM FOREST Model	17
Figure 5.3	XG Boost model	18
Figure 5.4	LSTM model	18
Figure 6.1	Onion Production in India	19
Figure 8.1	Missing value for Onion data	22
Figure 8.2	Top 10 Mandi for Onion based on Volume in metric tonnes	23
Figure 8.3	Top 10 Mandi for tomato based on Volume in metric tonnes	23
Figure 8.4	Top 10 Mandi for potato based on Volume in metric tonnes	23

Figure 8.5	Trend chart of Top 10 Mandi for Onion based on Retail Price in Rupees	24
Figure 8.6	Trend chart of Top 10 Mandi for Tomato based on Retail Price in Rupees	24
Figure 8.7	Trend chart of Top 10 Mandi for Potato based on Retail Price in Rupees	24
Figure 8.8	Pair plot showing data distribution for Bengaluru Onion Mandi data	25
Figure 8.9	Pair plot showing data distribution for Delhi Tomato Mandi data	25
Figure 8.10	Pair plot showing data distribution for Mumbai Potato Mandi data	25
Figure 8.11	Box plot for onion, tomato and potato retail prices	26
Figure 8.12	Anomalies for onion retail price in Bengaluru Mandi	26
Figure 8.13	Anomalies for tomato retail price in Delhi Mandi	27
Figure 8.14	Anomalies for potato retail price in Mumbai Mandi	27
Figure 9.1	seasonality trend Onion retail price in Bengaluru Mandi	32
Figure 9.2	seasonality trend tomato retail price in Delhi Mandi	32
Figure 9.3	seasonality trend Potato retail price in Mumbai Mandi	32
Figure 10.1	RMSE – evaluation results for onion	33
Figure 10.2	RMSE – evaluation results for tomato	34
Figure 10.3	RMSE – evaluation results for potato	34
Figure 11.1	Data pipe line deployment architecture proposal	35

List of Tables

No.	Name	Page No.
Table 6.1	Onion data sample	20
Table 6.2	Tomato data sample	20
Table 6.3	Potato data sample	21
Table 6.4	Diesel data	21
Table 9.1	p values for Onion retail price	29

Table 9.2	p values for tomato retail price	29
Table 9.3	p values for potato retail price	29
Table 9.4	ADF result for onion retail price	30
Table 9.5	ADF result for tomato retail price	30
Table 9.6	ADF result for potato retail price from Mumbai Mandi	31

Abstract

“Forecasting agricultural commodity futures prices is a crucial subject in agricultural domain, not only in providing price information of agricultural commodity in advance which decision makers believe, but also reducing the uncertainty and risks of agricultural markets” (Madaan et al., 2019)

“Fluctuations in commodity prices for onion, tomato and potato can cause distress among both consumers and producers, and are often exacerbated by trading networks especially in developing economies where marketplaces might not be operating under conditions of perfect competition for various contextual reasons” (Subhasree et al., 2016)

“India is mainly an agricultural country the farmer is an important part of agriculture. Agriculture mainly depends on him. Even then the farmers cannot predict prices for their commodities because prediction of prices plays a major challenge. Several characteristics are taken into account so that the crop price forecast is accurate. Forecasting price of agriculture commodities based on Volume, diesel price helps the agriculturist and also the agriculture mandi’s in India” (Varun et al., 2019)

“We look at onion, tomato and potato trading in India and present the evaluation of a price forecasting model, and an anomaly detection and compared different Supervised, Unsupervised and Forecasting prediction models. Our dataset consists of time series of wholesale prices, retail prices, arrival volumes and Diesel prices of the agricultural commodities at several mandi’s in India “ (Shome et al., 2018)

“I also provide an in-depth forecasting analysis of the effect on these retail price. Our results are encouraging and point towards the likelihood of building pricing models for agricultural commodities and to detect anomalies. These data can then be stored and analysed. the power to use historical data regarding agricultural commodities like onion, tomato and potato for the forecasting of retail prices are demonstrated” (Onion, 2016)

“I propose a comparative study of various forecasting strategies which will be wont to this aim. The empirical comparison of the chosen methods on the various data showed that some methods are more suitable than others for this type of problem. especially, we show that strategies supported Machine Learning approaches seem to be more suitable for this task” (Balaji Prabhu et al., 2018)

“I did a comparative study of Auto ARIMA (Autoregressive Integrated Moving Average), RNN (Recurrent Neural Network), LSTM, VAR (vector autoregressive model), Random forest Regression,XGBoost” (Li et al., 2019)

Keywords: Anomaly, commodity, Forecasting, Machine learning, Timeseries, Auto ARIMA, RNN, LSTM, VAR, Random forest Regression and XGBoost Regression

Contents

Candidate's Declaration 2
Certificate 3
List of Abbreviations 6
List of Figures 6
List of Tables 7
Abstract 9
Chapter 1: Introduction 11
Chapter 2: Literature Review 13
Chapter 3: Problem Statement 14
Chapter 4: Objectives of the Study 15
Chapter 5: Project Methodology 16
Chapter 6: Business Understanding 19
Chapter 7: Data Understanding 21
Chapter 8: Data Preparation 23
Chapter 9: Data Modeling 30
Chapter 10: Data Evaluation 36
Chapter 11: Deployment 38
Chapter 12: Analysis and Results 39
Chapter 13: Conclusions and Recommendations for future work 40
Bibliography 41
Appendix 44
- Plagiarism Report 44
- Publications in a Journal/Conference Presented/White Paper 47
- Any Additional Details 47

Chapter 1: Introduction

“Forecasting agricultural commodity futures prices will help to know the anomalies and other variables like diesel ,volume of onion, tomato and potato” (Madaan et al., 2019)

Data collected from different websites is knowledge for analysing past trends and commodity price patterns. Noisy and Nonstationary data means agricultural product prices can differ according to news and events. (Subhasree et al., 2016)

Noisy and Non stationary data lead to poor agricultural commodity price prediction results. “Forecasting prediction methods are generally based on the assumption that variables are independent, normal distribution, which is contradicted with real market” (Varun et al., 2019)

“Price fluctuations in agricultural commodities is an important area of study in forecasting. High prices increase the expenses of retail consumers while low prices reduce the incomes of farmer producers. In India, diesel is also one of the significant source of price variation since the majority of agricultural production is transported from different places to commodity mandi’s” (Shome et al., 2018)

The prices of agricultural commodities for onion, tomato and potato are volatile. Noisy and Non stationary data and events affect onion, tomato and potato retail prices.

“The prices of agricultural commodities are forecasted using various time-series, machine learning, and deep learning models. Onion, tomato and potato is an important commercial tree crop mainly cultivated in Mumbai, Bengaluru and Delhi. Although there exist numerous online platforms such as commodity online, that provides details of prices of onion, tomato and potato from various cities predicted using VAR, AutoARIMA, LSTM,Random forest and XGBoost. The performance of these models was evaluated and compared to identify the best model” (Zhang et al., 2020)

“The work focuses on analysing such agricultural data using techniques of machine learning and gaining the knowledge obtained from the result. The comparison of results obtained from several algorithms will help in selecting the more appropriate algorithm for agricultural data to predict retail prices of onion, tomato and potato from 2015 to 2020” (Ouyang et al., 2019)

Agriculture commodity Wholesale price, retail price and volume for onion, tomato and potato has been a collected from Horticulture website from 2015 – 2020. Diesel Price also collected from 2015-2020 and it contributes in deriving retail price of the agricultural commodity. During this statistical study of the data we observed several anomalies which are explained with more details in the chapter 12. (Istambul et al., 2019)

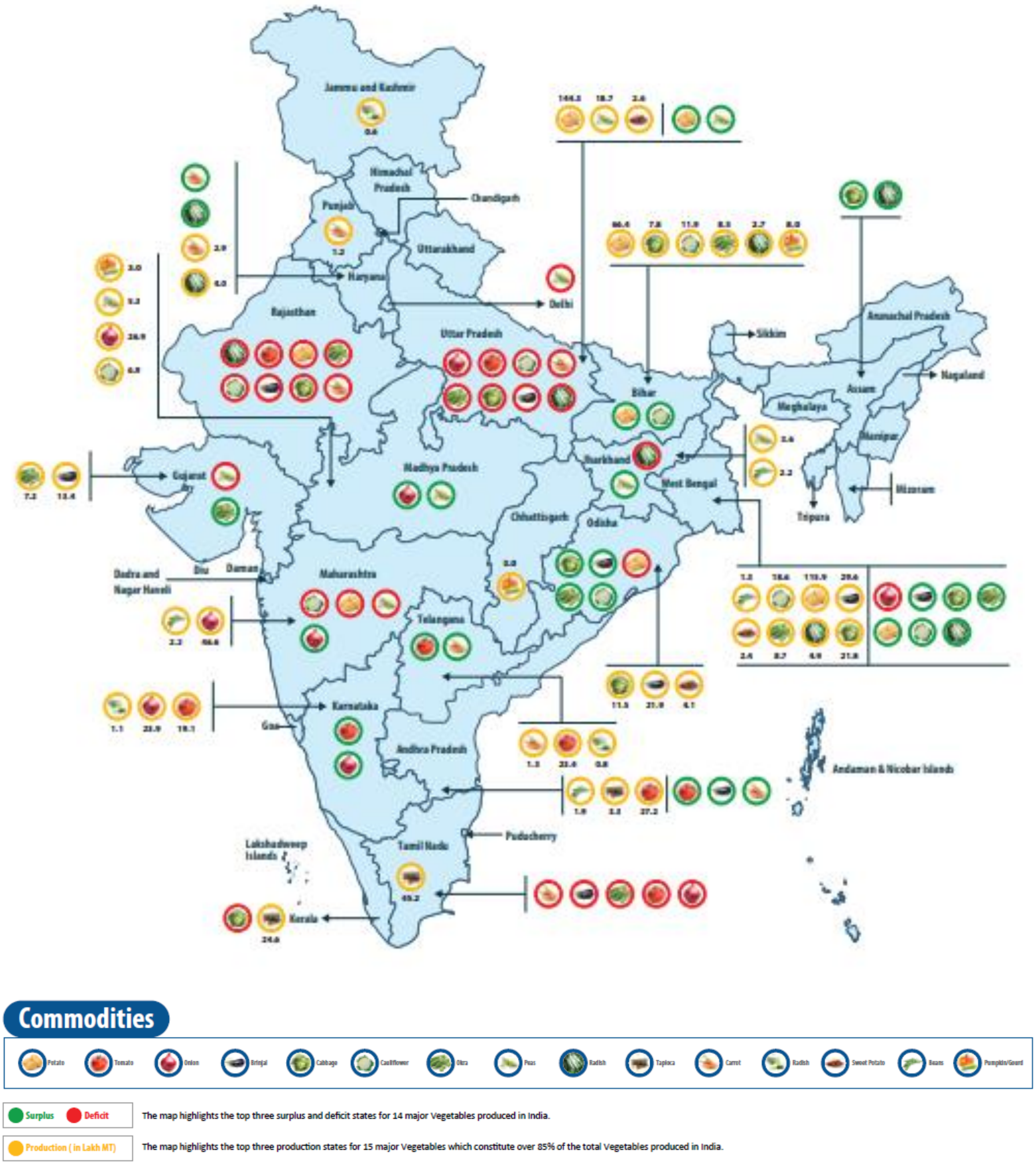

Figure 1.1: Agricultural commodity of India, (India, 2020)

Chapter 2: Literature Review

In this project we study several mandi's across the India for onion, tomato based on arrival volumes and diesel prices between 2015 to July 2020. Identified top mandi for onion, tomato and potato for the prediction of retail price.

The wholesale, retail, volume and diesel price data collected from different websites available online.(Agmarket, 2020) (Gati, 2020)(India, 2020)

Diesel price also play a major role in deciding retail prices of the agricultural commodity, Diesel price change every day due to changes in international crude oil price and policies. (Shome et al., 2018)

"Many researchers have implemented various models for forecasting agricultural commodity prices and their determinants. Most empirical studies on forecasting agricultural commodity prices rely on econometric models or intelligent algorithms" (Onion, 2016)

Data which are noisy represent less information in the collected data to look from the historical pattern of agricultural commodity prices of onion, tomato and potato. Data which are non-stationary represent onion, tomato and potato prices may change in different timeseries due to different events. (Ouyang et al., 2019)

Research work done to identify anomalies in the prediction of onion, tomato and potato retail price. In our project we identified the anomalies based on the events which results in sudden rise in price trend. (Onion, 2016) (Madaan et al., 2019)

"It is a great opportunity to implement and compare different machine learning algorithms with the assumption of variables are not dependent, with normal distribution and with real time data collected from different websites. In our project I did a comparison of different timeseries forecasting models with supervised/unsupervised prediction models to predict the retail price" (Bratsas et al., 2020)

There are different studies performed prediction of deep learning – LSTM and other methods which will be helpful to handle large amount of data using TensorFlow, keras in Python. (Ouyang et al., 2019)

Work done in comparing different performance result RMSE for all the forecasting prediction models will help to choose the right models for the prediction of onion, tomato and potato retail price. These are key evaluating parameters to identify best prediction models for retail price of onion, tomato and potato. (Subedar et al., 2017)

Chapter 3: Problem Statement

The problem statement of the "Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato" is to develop a machine learning models that can accurately predict the retail price of onion, tomato and potato from the collected data.

The "National Horticulture Board runs a portal which provides retail prices from across 30 district centers across the country" (India, 2020). I did web crawling to collect data from January 2015 till July 2020. I also collected diesel prices in India between January 2015 till July 2020.

"The collected data from different websites is defined as the set of features selected for the agricultural commodity in India – onion, tomato and potato" (India, 2020)

The idea behind the problem is that an accurate forecasting model will be able to reduce the variation in forecasting of retail prices successfully.

During the study of different forecasting prediction models, also need to identify different anomalies which are outside the range of 2 sigma levels.

Need to identify key events based on anomalies which resulted in rise or fall or volume and resulted in changes in retail prices of the agricultural commodities – onion, tomato and potato.

This problem is also a supervised task because the targets for the training data are known future of time and the model will learn based on tabular data.

Chapter 4: Objectives of the Study

“Onion, tomato and potato commodity retail prices are influenced by a combination of factors, arrival volume, diesel price, anomalies etc. These factors cannot be quantified by the same scale, and have different effect on onion, tomato and potato in different wholesale markets in India” (Madaan et al., 2019). The prediction of monthly forecasting is challenging due to changes of prices is affected by a combination of different factors

In this project we are focusing on following “types of short-term forecasting methods to predict the agricultural commodity prices of onion, tomato and potato”(Balaji Prabhu et al., 2018)

“Time series methods, including short-term forecasting methods like Auto ARIMA model. These approaches are based on agricultural product historical prices only, thus ignoring other variables. These models no longer work when non-stationary factors influence prices” (Subedar et al., 2017)

Regression methods, including vector auto-regression model, Random forest, XG Boost. These methods take other factors into consideration. However, because of the constraint of the conditions of use, it is difficult to use a single model at the same time to forecast many different types of agricultural commodities (Balaji Prabhu et al., 2018)

Deep learning methods, including neural networks, LSTM also studied to compare it with other forecasting models. “These methods have extensive application scope. However, when forecasting different agricultural commodities, the effects cannot be ensured and overfitting may happen. Thus, these methods are usually used to forecast some specific kinds of agricultural commodities. The method proposed in this project has good forecasting effects on onion, tomato and potato retail prices in Indian agricultural markets”(Arunraj et al., 2016) (Zhang et al., 2020)

The main objective of this project basically includes:

- The development of the model to forecast and predict a monthly price to calculate and forecast the pattern of monthly changes in agricultural commodities.
- The ML models can be used to identify the monthly prices of onion, tomato and potato agricultural commodities retail prices with good effects.
- Comparative study done to know which are the suitable models we can implement to forecast real time data to predict onion, tomato and potato prices.
- Identify the anomalies from the data.

Chapter 5: Project Methodology

The Methodology followed in this project is The "**CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (*CRISP-DM*) is a process model with six phases that naturally describes the data science life cycle:

- Business understanding
- Data understanding
- Data preparation
- Modelling
- Evaluation
- Deployment" (Wikipedia, 2020a)

Figure 5.1: CRISP-DM lifecycle
(Wikipedia, 2020a)

In this project we worked on the Auto ARIMA, VAR, Random forest, XG Boost, LSTM to predict retail prices of onion, tomato and potato.

Auto ARIMA – is a method of supervised learning for issues relating to time series. Auto ARIMA is essentially ARIMA with a gridsearch part to find the best metric parameters such as the Akaike Knowledge Criterion (AIC). ARIMA stands for Autoregressive Integrated Moving Average. ARIMA can also be broken down into part Autoregressive (AR), part Moving Average (MA) and part Integrated (I). Models AR and MA can be used as separate models to produce predictions of the time series. It can however produce better predi ctions when combined as an ARMA model.

“VAR - Vector autoregression (VAR) is a statistical model used to capture multiple quantity relationships as they change over time. VAR is a model of the stochastic process. The VAR models generalise the autoregressive single variable (univariate) model by allowing multivariate time series” (Wikipedia, 2020e)

“The model of vector autoregression (VAR) adds the idea of univariate autoregression to kk time series regressions, in which all kk series lagged values appear as regressors. Put differently, a vector of time series variables on lagged vectors of those variables are regre ssed in a VAR model. As for AR (pp) models, the lag order is denoted by pp so the VAR(pp) model of two variables Xt and Yt (k=2k=2) is given by the equations” (Wikipedia, 2020e)

“Yt=β10+β11Yt−1+⋯+β1pYt−p+γ11Xt−1+⋯+γ1pXt−p+u1t,
Xt=β20+β21Yt−1+⋯+β2pYt−p+γ21Xt−1+⋯+γ2pXt−p+u2t” (Wikipedia, 2020e)

“The ββs and γγs can be estimated using OLS on each equation” (Wikipedia, 2020e)

“Random forest - The "forest" it builds, is an ensemble of decision trees , usually trained by the method of "bagging." The general idea of the bagging method is that the overall result increases with a combination of learning models. Advantages of Random Forest algorithm – it will avoid the overfitting problem, handle missing values identify important features and can be used in both classification and regression problems” (Zhang et al., 2020)

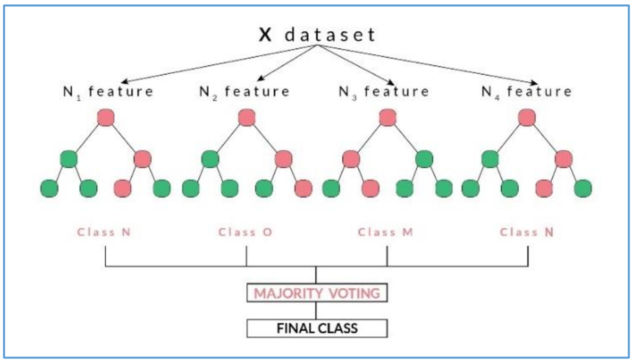

Figure 5.2: Random forest model (Ali et al., 2012)

“XGBoost - is a Machine Learning algorithm based on a decision- tree ensemble that uses a gradient boosting system” (Timothy Susanto, 2020)

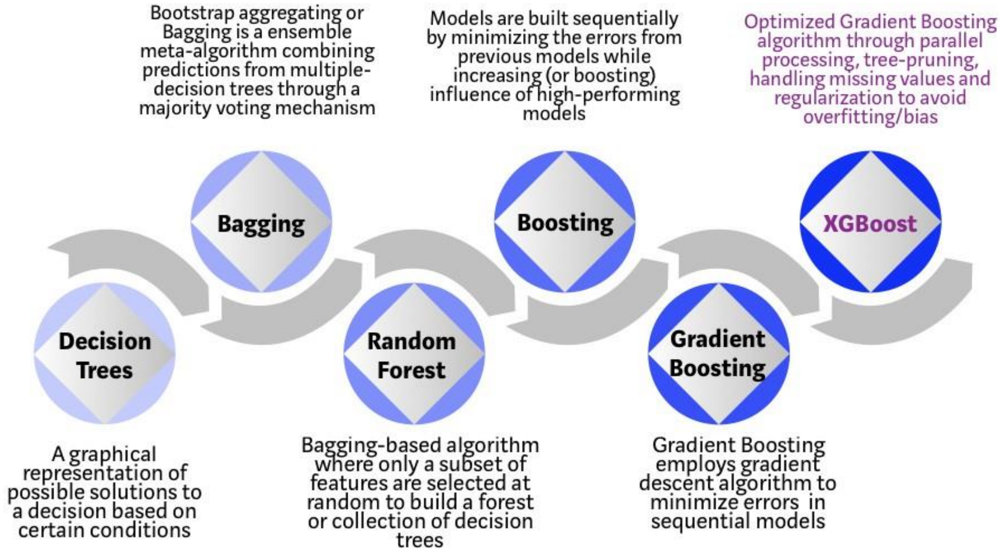

Figure 5.3: XGBoost model (Timothy Susanto, 2020)

“ LSTM - Long Short-Term Memory networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory” (Madaan et al., 2019)

“LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation. In an LSTM network, three gates are present: Input gate, output gate and forget gate” (Madaan et al., 2019)

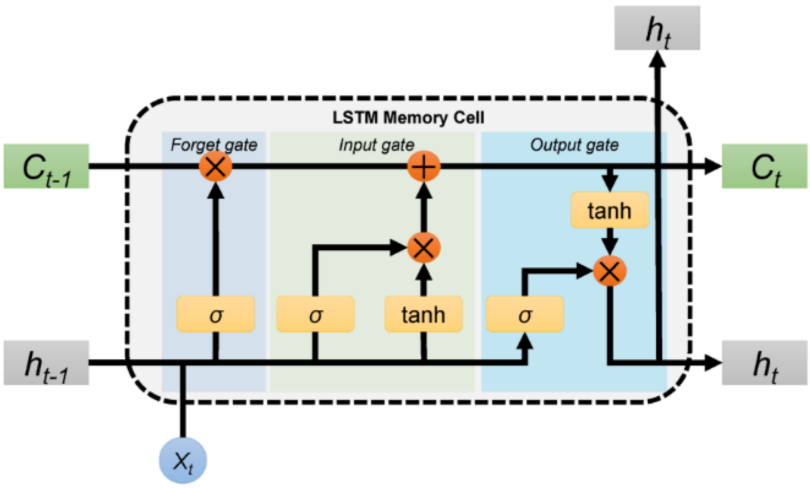

Figure 5.4: LSTM model (Madaan et al., 2019)

Chapter 6: Business Understanding

"India ranks second worldwide in farm outputs. As per 2018, agriculture employed more than 50% of the Indian work forces and contributed 17–18% to country's GDP. According to latest report, agriculture is primary source of livelihood for 58% population in India" (Wikipedia, 2020b)

"Onion (Allium cepa) is one of the second most important commercial crops of the India which is next to Potato.

In the world, Onion crop is grown in about 5.30-million-hectare area with an annual production of 88.48 million tons with productivity 16.70 tons per hectare.

In India, Onion crop is grown in about 1.20-million-hectare area with an annual production of 19.40 million tons with productivity 16.12 tons per hectare.

The quantity of Onion 2415.75 thousand tons is exported from India which outputs value of 3,10,650.09 Rs. Lakhs (in 2017)" (Wikipedia, 2020b)

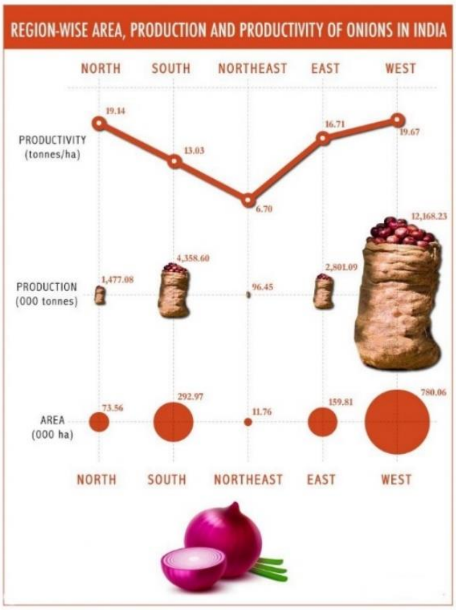

Figure 6.1: Onion Production in India (Murthy, 2019)

"Tomato is one of the most important vegetable crops cultivated for its fleshy fruits. Tomato is considered as important commercial and dietary vegetable crop. Botanical name of tomato is Lycopersicon esculemtun and belongs to family Lycopersicae. As it is short duration crop and gives high yield, it is important from economic point of view and hence area under its cultivation is increasing day by day. Tomato is used in preserved products like ketch-up, sauce, chutney, soup, paste, puree etc. The estimated area and production of tomato for India are about 3,50,000 hectares and 53,00,000 tons respectively" (Wikipedia, 2020d)

“Potato (Solanum tuberosum) is the most important food crop of the world. Potato is a temperate crop grown under subtropical conditions in India. India’s potato production is 52.5 million tons in 21.8 lakh hectares” (Wikipedia, 2020c)

I collected data from horticultural website of India for onion, tomato and potato, as we know onion, tomato and potato are major agricultural crops in India. Our study will help to predict onion, tomato and potato retail prices.

Chapter 7: Data Understanding

"The National horticultural board website run by the Government of India makes publicly available the monthly data on mandi wholesale and retail prices in Rupees and arrival volume in Metric tonne of onion, tomato and potato from different mandi's in India. We scraped all the data for onions, tomato & potatoes from all mandis for 5 years from January 1st 2015 to July 2020" (India, 2020)

Diesel price in rupees per litre also collected from website (Gati, 2020) which is also contributing in prediction of retail prices, we collected for 5 years from January 1st 2015 to July 2020.

There were missing values in all these commodities which will be treated as explained in Chapter 8.

<u>Dataset for Onion commodity (sample from full data):</u>

Date	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice	AHMEDABAD_Volume	Bengaluru_WholesalePrice	Bengaluru_RetailPrice	Bengaluru_Volume
1/1/2015	1359	2317	13405	1758	2791	74978
2/1/2015	1569	2300	13009	1815	2750	42164
3/1/2015	1389	2285	11983	1576	2396	41339
4/1/2015	1099	2300	14303	1455	2260	43534
5/1/2015	1140	2300	13349	1725	2458	44227
6/1/2015	1582	2486	11567	2182	3145	43403
7/1/2015	1957	3375	11451	2357	3314	45957
8/1/2015	3838	5571	7103	2727	4029	83437
9/1/2015	4180	6762	7425	2536	4391	114533
10/1/2015	2757	5958	13974	2223	3914	173788
11/1/2015	1881	4023	17691	2664	4264	77516
12/1/2015	1283	2889	17078	1494	3000	53853

Table 6.1: Onion data sample in India (India, 2020)

<u>Dataset for Tomato commodity (sample from full data):</u>

Date	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice	AHMEDABAD_Volume	Bengaluru_WholesalePrice	Bengaluru_RetailPrice	Bengaluru_Volume
1/1/2015	1194	3125	3658	1295	1961	30762
2/1/2015	1198	3205	3318	829	1392	11995
3/1/2015	880	2600	2820	776	1296	13511
4/1/2015	895	2477	3652	995	1640	8964
5/1/2015	1231	3190	6883	2115	2888	11665
6/1/2015	1402	3182	7446	1318	2095	11969
7/1/2015	1904	3771	8094	1748	2505	10460
8/1/2015	1114	3071	8125	631	1204	13605
9/1/2015	1124	3071	8440	705	1318	12390
10/1/2015	1360	3146	5635	1464	2205	13393
11/1/2015	2275	3977	3339	3236	4464	10978
12/1/2015	1194	2889	2286	2364	3616	8097

Table 6.2: Tomato data sample in India nhb (India, 2020)

<u>Dataset for Potato commodity (sample from full data):</u>

Date	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice	AHMEDABAD_Volume	Bengaluru_WholesalePrice	Bengaluru_RetailPrice	Bengaluru_Volume
1/1/2015	815.00	2450.00	22231.00	1758.00	2770.00	32806.00
2/1/2015	576.00	2000.00	17116.00	1569.00	2592.00	21241.00
3/1/2015	348.00	1550.00	22966.00	1246.00	2274.00	22587.00
4/1/2015	318.00	1473.00	7729.00	1153.00	2105.00	22022.00
5/1/2015	381.00	1463.00	8255.00	1004.00	2050.00	23161.00
6/1/2015	513.00	1500.00	5256.00	1048.00	2091.00	22649.00
7/1/2015	419.00	1500.00	9134.00	929.00	2000.00	21285.00
8/1/2015	434.00	1895.00	10844.00	863.00	2000.00	27800.00
9/1/2015	482.00	2000.00	9793.00	786.00	2000.00	23551.00
10/1/2015	601.00	2000.00	9578.00	805.00	2027.00	28165.00
11/1/2015	679.00	2000.00	3368.00	1086.00	2145.00	25731.00
12/1/2015	963.00	2000.00	3451.00	910.00	2040.00	29308.00

Table 6.3: Potato data sample in India (India, 2020)

Date	DieselPrice
1/1/2015	53.74
2/1/2015	51.96
3/1/2015	55.41
4/1/2015	52.56
5/1/2015	58.19
6/1/2015	56.69
7/1/2015	53.37
8/1/2015	48.23
9/1/2015	47.7
10/1/2015	49.26
11/1/2015	53.11
12/1/2015	49.21

Table 6.4: Diesel price in Rupees data sample in India (Gati, 2020)

<u>Data Understanding Summary:</u>

- Downloaded **onion,** tomato **and potato** data from January 2015 till July 2020 from National horticultural (India, 2020) and Diesel price from gatti website (Gati, 2020)

- Combined data for commodity prices, volume and diesel prices into onion, tomato and potato using panda's data frames, respectively
- Data is having missing values <15%, total 67 rows of data and for different mandis for onion, tomato and potato respectively.

Chapter 8: Data Preparation

Following are the activities during data preparation:

- Missing value Imputation
- Identify Top 10 mandis of India for Onion, tomato and potato based on volume
- Data Distribution, Scaling
- Exploratory data analysis

Missing value Imputation: Observed missing value in onion, tomato, potato and diesel data. We did missing value imputation using replacing missing value by mean.

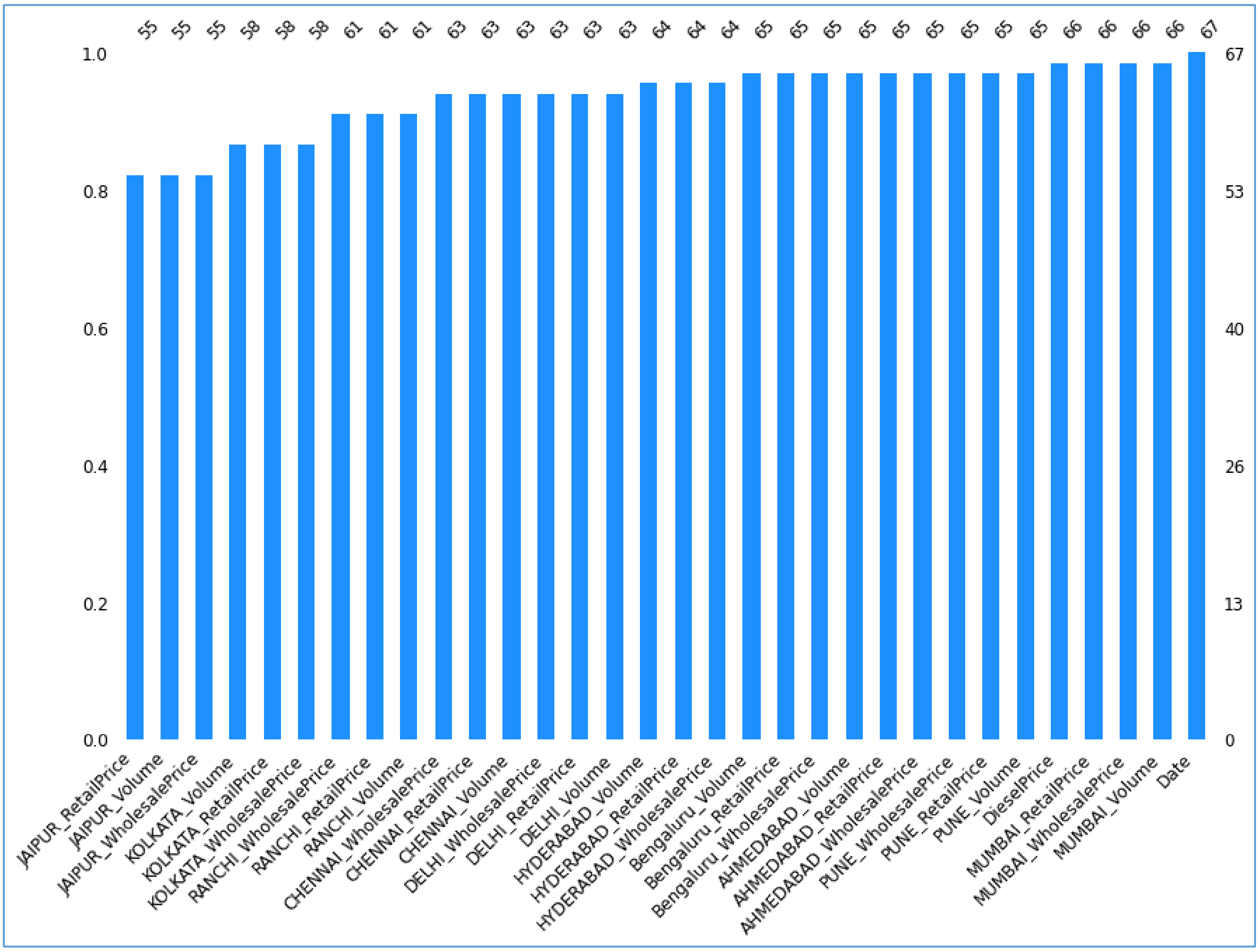

Figure 8.1: Missing value for Onion data

<u>Identify of Top 10 mandis of India for Onion, tomato and potato based on volume:</u>

Identified top 10 mandis of India for onion, tomato and potato considering overall volume. From this information, I chose top mandi for onion, tomato and potato to do further analysis and models.

Top mandis for onion is Bengaluru, Tomato is Delhi and for potato is Mumbai.

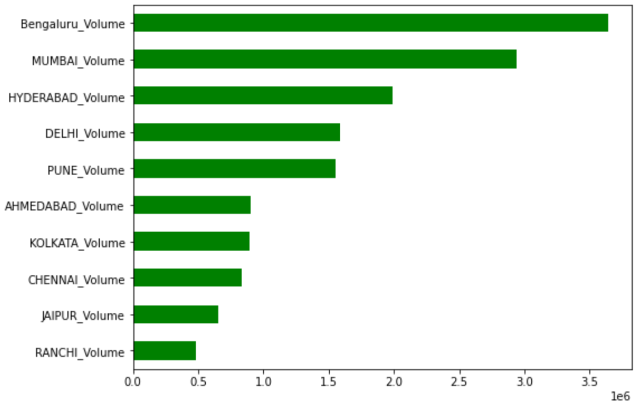

Figure 8.2: Top 10 Mandi for Onion based on Volume in metric tonnes

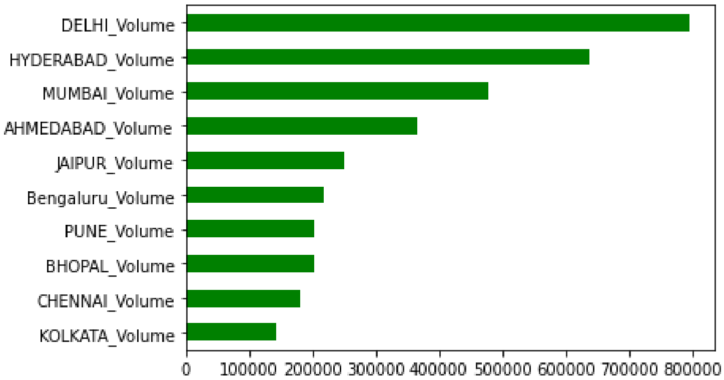

Figure 8.3: Top 10 Mandi for Tomato based on Volume in metric tonnes

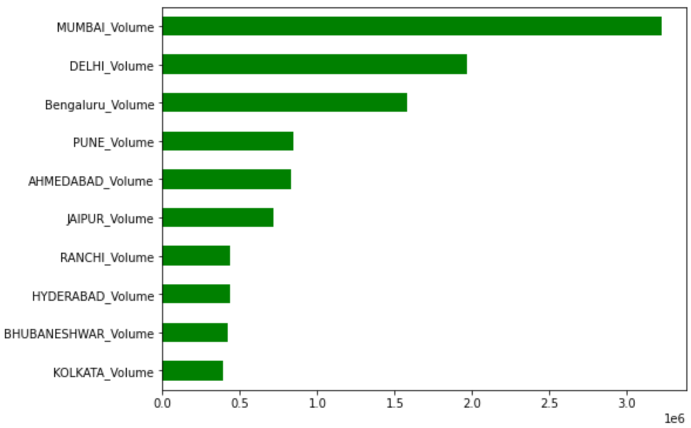

Figure 8.4: Top 10 Mandi for Potato based on Volume in metric tonnes

The following trend chart shows the retail price trend for 10 mandis of onion, tomato and potato from January 2015 till July 2020.

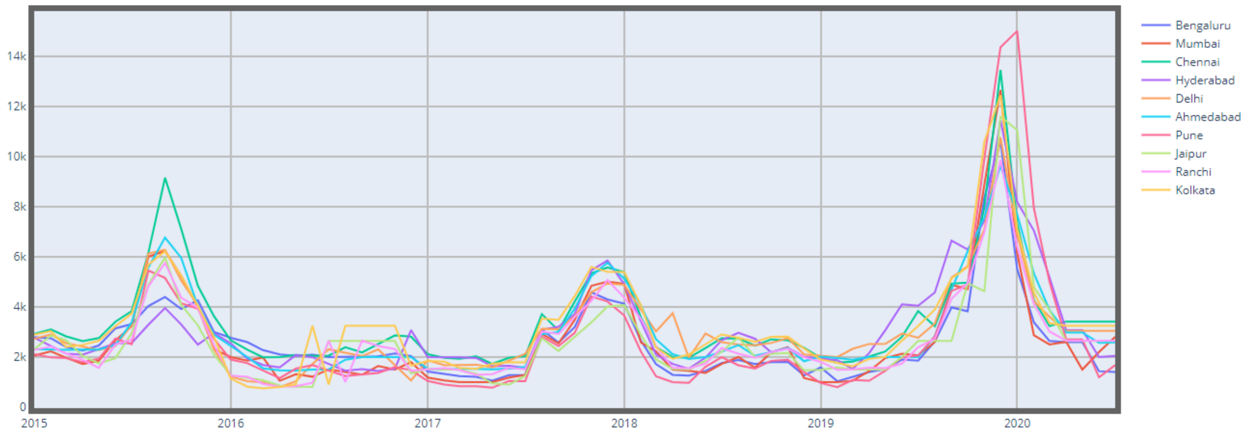

Figure 8.5: Trend chart of Top 10 Mandi for Onion based on Retail Price in Rupees

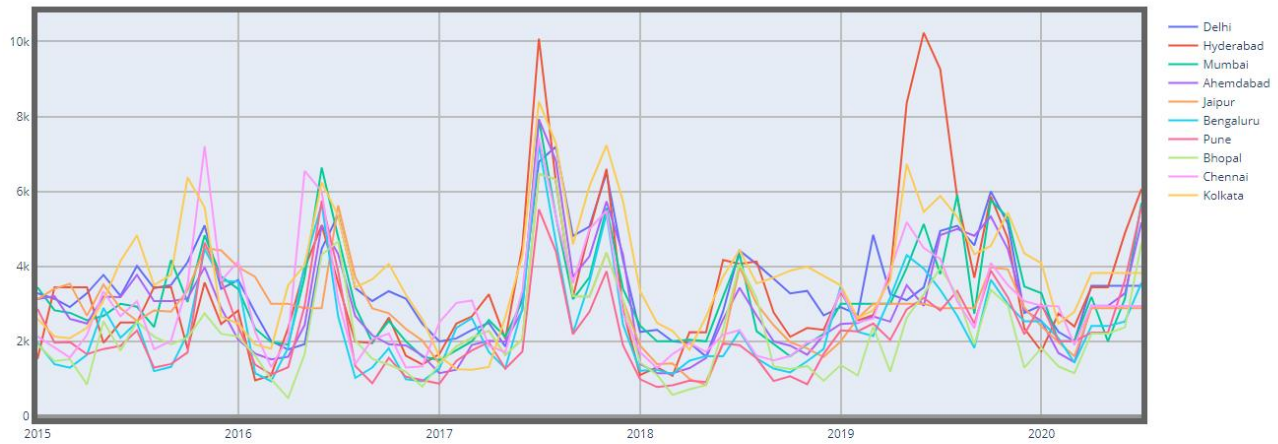

Figure 8.6: Trend chart of Top 10 Mandi for Tomato based on Retail Price in Rupees

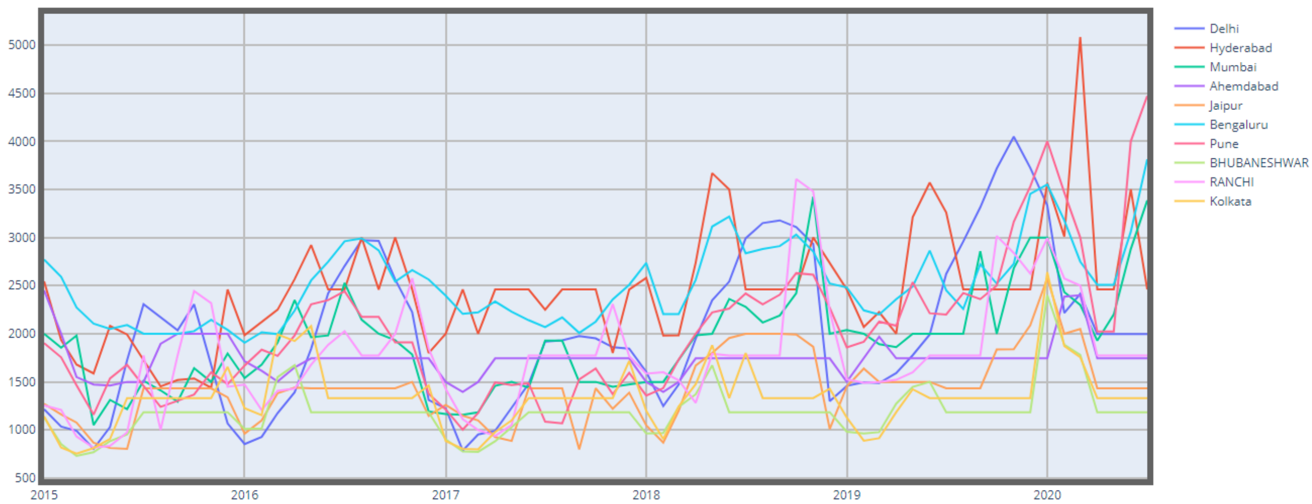

Figure 8.7: Trend chart of Top 10 Mandi for Potato based on Retail Price in Rupees

Following pair plot show the data distribution and correlation between each of the variables from one of the Top mandis for onion, tomato and potato

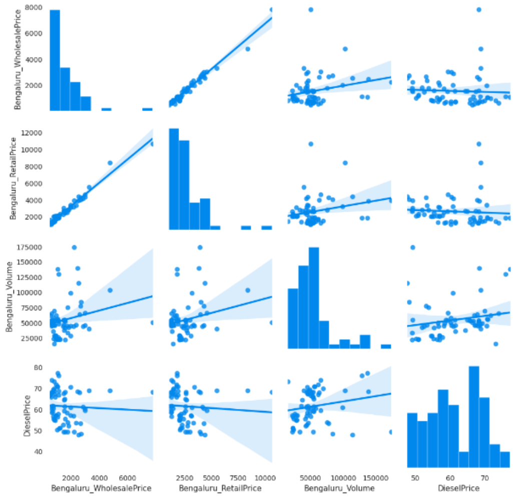

Figure 8.8: Pair plot showing data distribution for Bengaluru Onion Mandi data

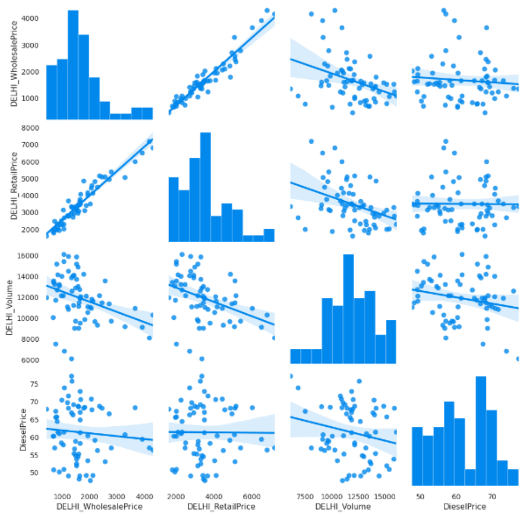

Figure 8.9: Pair plot showing data distribution for Delhi Tomato Mandi data

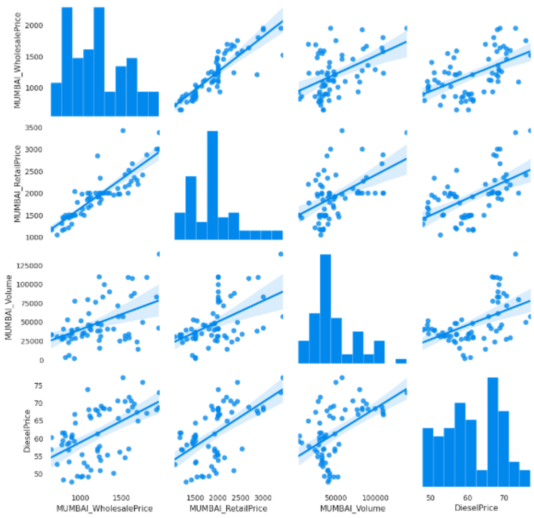

Figure 8.10: Pair plot showing data distribution for Mumbai Potato Mandi data

Identified outliers in retail price for onion, tomato and potato.

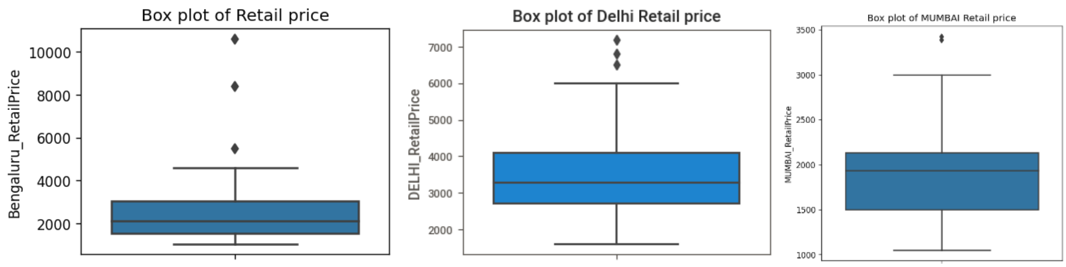

Figure 8.11: Box plot for onion, tomato and potato retail prices

<u>Data Scaling:</u> Variables calculated at various scales do not contribute equally to the model fitting & model role learned and which ultimately generate a bias.

Thus, standardised feature-wise (μ=0, σ=1) is typically used before model fitting to solve this potential issue. StandardScaler does away with the mean and measures the variance any features. StandardScaler eliminates the mean and scales to unit variance for each function / variable. This scaling is achieved independently of feature-wise.
I used Scikit python 's basic scaler function to scale variables for onion, tomato and potato.

<u>"Anomalies:</u> Unexpected values often surface in a distribution of values, especially when working with data from unknown sources which lack poor data validation controls" (Bhattacharya, 2020).

Detailed explanation of these anomalies will be in chapter 12.

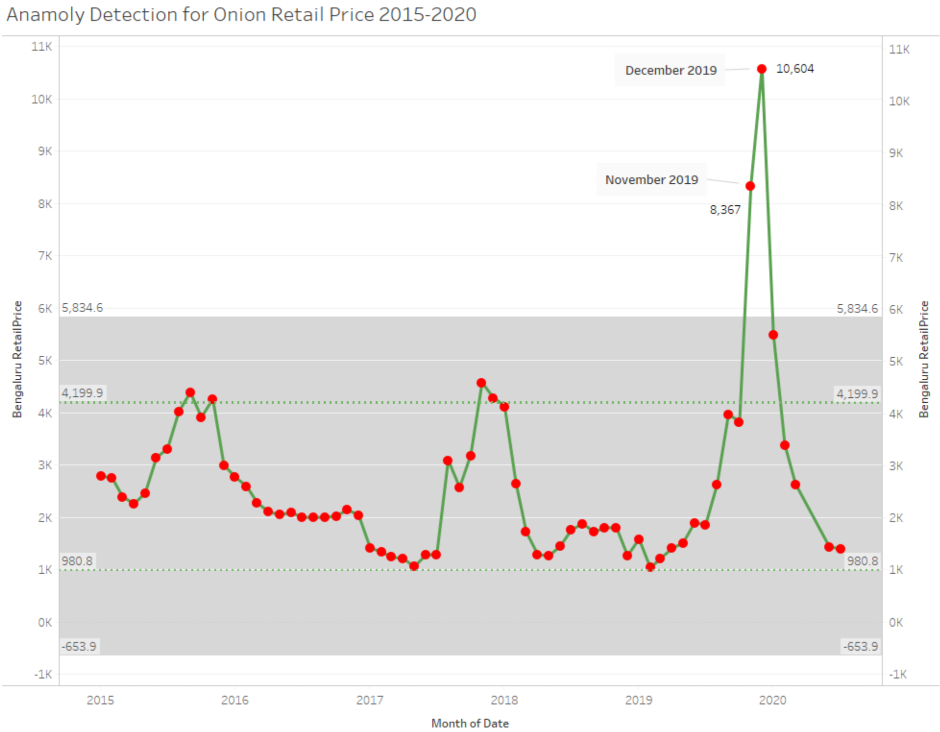

Figure 8.12: Anomalies for onion retail price in Bengaluru Mandi

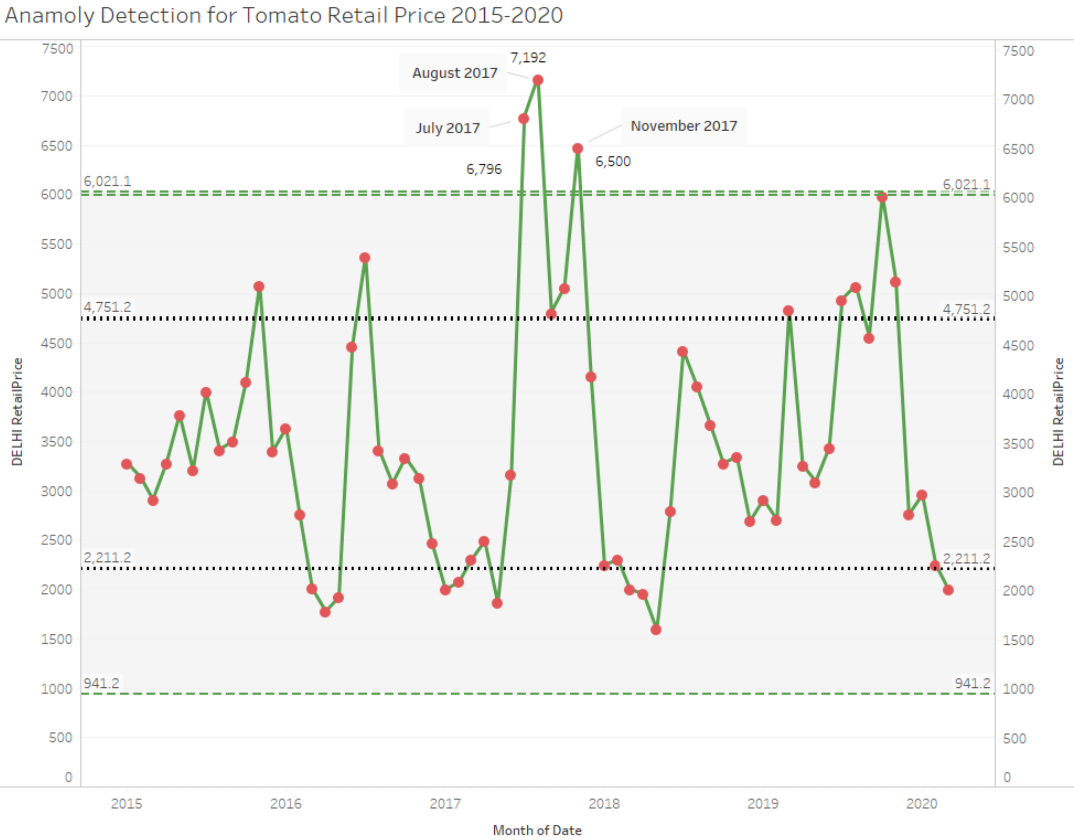

Figure 8.13: Anomalies for tomato retail price in Delhi Mandi

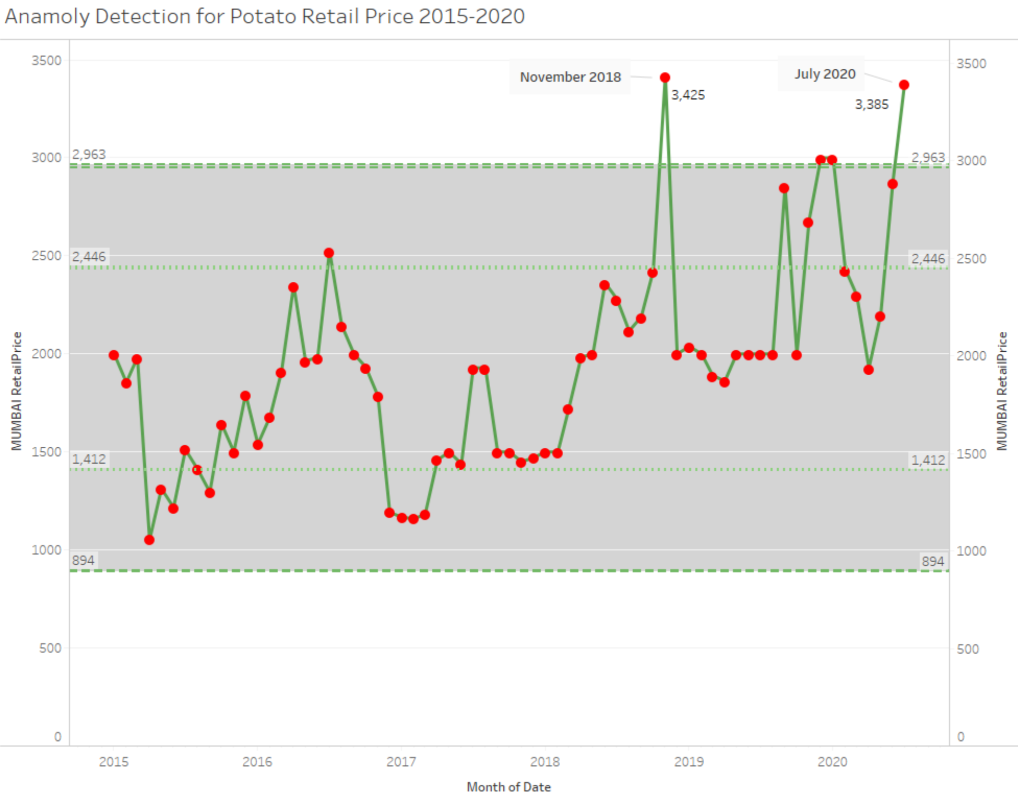

Figure 8.14: Anomalies for potato retail price in Mumbai Mandi

Data Preparation Summary:

Performed the following activities and feature engineering steps:

Data Cleaning

Identified top 10 mandi's for onion, tomato and potato

Missing value imputation for all the missing value by mean

Feature Engineering

Understanding of data distribution

Understanding of trend for Retail prices of onion, potato and tomato

Identified anomalies and outliers in the top mandi's – Bengaluru, Delhi and Mumbai for onion, tomato and potato respectively

Output

Clean dataset for Bengaluru (Onion), Delhi (Tomato) and Potato (Mumbai) exported to do data modeling and further machine learning activities.

Chapter 9: Data Modeling

Modeling is performed by splitting train and test for Onion, tomato and potato data from Bengaluru, Delhi and Mumbai Mandi's respectively.

In our datasets splitting of train and test is done in the ratio of 75%, 25% respectively.

The train set contains an output from the given data and the models we built learns using this data. I also have the test data in order to test my model's prediction performance. RMSE (root mean square error) is the evaluation metrics used to choose best model for the retail price prediction.

Following statistical test performed on onion, tomato and potato for Bengaluru, Delhi and Mumbai mandi's are as follows:

9.1 Testing Causation using Granger's Causality Test

9.2 Cointegration Test

9.3 Augmented Dickey-Fuller Test (ADF Test)

9.4 Durbin Watson test

9.5 Auto correlation and seasonality function analysis

9.1 Testing causation using Granger's causality: "Granger causality is a method of studying the causality in a time series between two variables. The approach is a probabilistic account of causality, applies to data to identify correlation patterns"(Anil Seth, 2007). "The values in the table are the P-Values. P- value lesser than the significance level (0.05), implies the Null Hypothesis that the coefficients of the corresponding past values is zero, that is, the X does not cause Y can be rejected" (Anil Seth, 2007)

	Bengaluru_RetailPrice_x	Bengaluru_Volume_x	DieselPrice_x
Bengaluru_RetailPrice_y	1.0000	0.0001	0.0004
Bengaluru_Volume_y	0.0236	1.0000	0.0026
DieselPrice_y	0.0015	0.0351	1.0000

Table 9.1: p values for Onion retail price from Bengaluru Mandi

	DELHI_RetailPrice_x	DELHI_Volume_x	DieselPrice_x
DELHI_RetailPrice_y	1.0000	0.0140	0.1208
DELHI_Volume_y	0.1466	1.0000	0.2293
DieselPrice_y	0.0368	0.0001	1.0000

Table 9.2: p values for Tomato retail price from Delhi Mandi

	MUMBAI_RetailPrice_x	MUMBAI_Volume_x	DieselPrice_x
MUMBAI_RetailPrice_y	1.0000	0.0001	0.0499
MUMBAI_Volume_y	0.0081	1.0000	0.0008
DieselPrice_y	0.1412	0.0234	1.0000

Table 9.3: p values for Potato retail price from Mumbai Mandi

9.2 "Cointegration Test: helps to establish the presence of a statistically significant connection between two or more time series. This we have performed using Johansen test.

The results show retail prices of onion, potato and tomato is having significant connection with time series"(Corporate finance institute, 2020)

9.4 **"Augmented Dickey-Fuller Test (ADF Test):** The Augmented Dickey Fuller Test (ADF) is Stationary Unit Root Test" (wiki, 2020). Unit roots in your analyses of the time series will trigger unpredictable results. These test are run for Bengaluru, Mumbai and delhi mandi data for onion, tomato and potato to understand the data is stationary or not. The test is run twice before and after differencing to ensure data is stationary and rejecting null hypothesis.

```
   Augmented Dickey-Fuller Test on "Bengaluru_RetailPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -7.3321
No. Lags Chosen       = 1
Critical value 1%     = -3.581
Critical value 5%     = -2.927
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "Bengaluru_Volume"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -5.8933
No. Lags Chosen       = 3
Critical value 1%     = -3.589
Critical value 5%     = -2.93
Critical value 10%    = -2.603
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "DieselPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -8.5255
No. Lags Chosen       = 2
Critical value 1%     = -3.585
Critical value 5%     = -2.928
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.
```

Table 9.4: ADF result for onion retail price from Bengaluru Mandi

```
   Augmented Dickey-Fuller Test on "DELHI_RetailPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -10.2258
No. Lags Chosen       = 1
Critical value 1%     = -3.581
Critical value 5%     = -2.927
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "DELHI_Volume"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -3.6628
No. Lags Chosen       = 10
Critical value 1%     = -3.621
Critical value 5%     = -2.944
Critical value 10%    = -2.61
=> P-Value = 0.0047. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "DieselPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -8.5255
No. Lags Chosen       = 2
Critical value 1%     = -3.585
Critical value 5%     = -2.928
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.
```

Table 9.5: ADF result for tomato retail price from Delhi Mandi

```
   Augmented Dickey-Fuller Test on "MUMBAI_RetailPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -7.6056
No. Lags Chosen       = 2
Critical value 1%     = -3.585
Critical value 5%     = -2.928
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "MUMBAI_Volume"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -5.8797
No. Lags Chosen       = 5
Critical value 1%     = -3.597
Critical value 5%     = -2.933
Critical value 10%    = -2.605
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.


   Augmented Dickey-Fuller Test on "DieselPrice"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -8.5255
No. Lags Chosen       = 2
Critical value 1%     = -3.585
Critical value 5%     = -2.928
Critical value 10%    = -2.602
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.
```

Table 9.6: ADF result for Potato retail price from Mumbai Mandi

9.4 Durbin Watson test: Is a test of autocorrelation (also known as serial correlation) / tests for homoscedasticity in regression residuals. This test also can be performed using OLS (ordinary least square) method.

Autocorrelation is a time series similitude over consecutive periods of time. It can cause the standard error to be underestimated, and can cause predictorsto be large when not. This test searches for a particular form of serial correlation, the AR (1) method.

Thus, the test statistic equals 2 for r = 0, suggesting no serial correlation. This number often r anges from 0 to 4. The closer the numbers are to 0, the greater the proof for a strong serial cor relation. The closer to the four, the greater the evidence of negative serial correlation.

```
Bengaluru_RetailPrice : 1.61
Bengaluru_Volume : 1.6
DieselPrice : 0.82
```

```
DELHI_RetailPrice : 1.5
DELHI_Volume : 1.62
DieselPrice : 2.87
```

```
MUMBAI_RetailPrice : 1.4
MUMBAI_Volume : 0.6
DieselPrice : 1.11
```

9.5 Auto correlation function (ACF) and seasonality analysis: ACF is a form of dependency on sequence. Autocorrelation, in particular, is when a time series is linearly conn ected to a lagged version of itself. Partial autocorrelation solves this problem by measuring the correlation between variables and it’s timeseries when the influence of the intermediate variables has been removed. In comparison, correlation is simply when there are linear ties between two independent variables, the trend in time series data, deleted the seasonality.

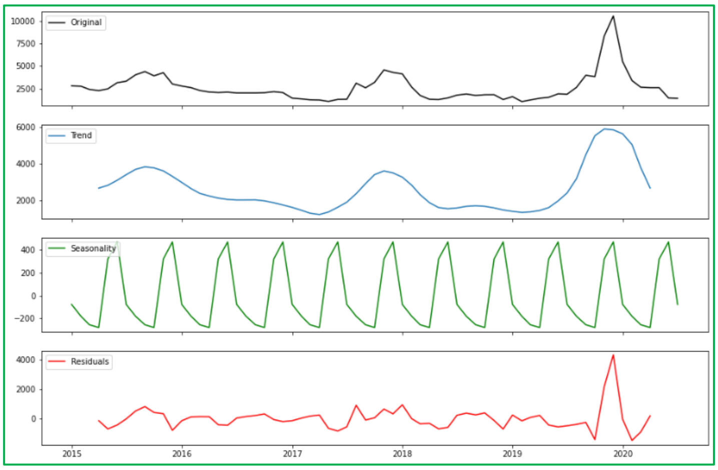

Figure 9.1: seasonality trend onion retail price in Bengaluru Mandi

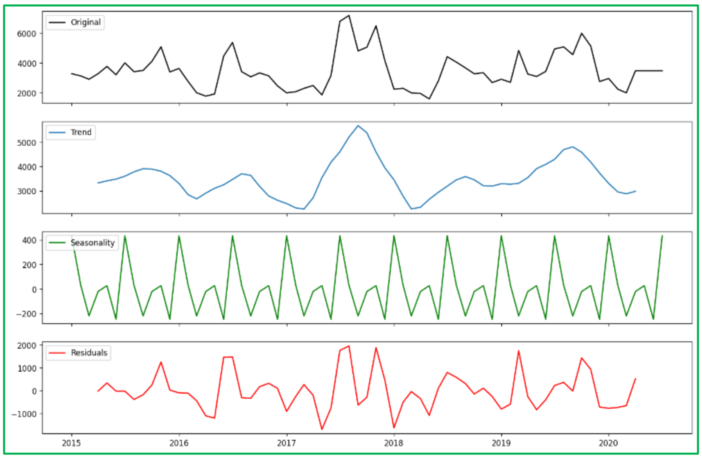

Figure 9.2: seasonality trend tomato retail price in Delhi Mandi

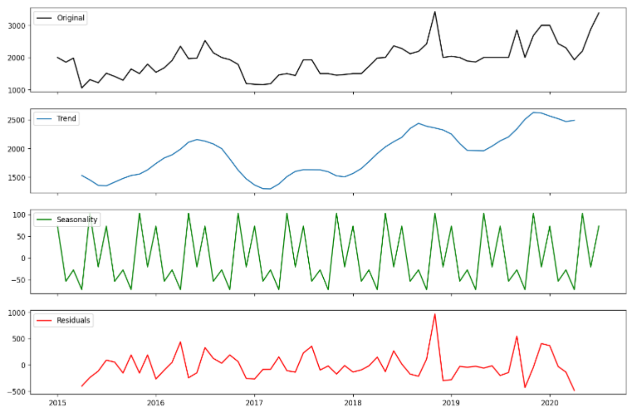

Figure 9.3: seasonality trend Potato retail price in Mumbai Mandi

RMSE – Root Mean Square Error is a standard method for calculating a model 's error when predicting quantitative results. It is formally defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n \text{ are predicted values}$$
$$y_1, y_2, \ldots, y_n \text{ are observed values}$$
$$n \text{ is the number of observations}$$

Source: (WIKI, 2020)

RMSE has a twofold aim: To act as a heuristic model for the training,
to assess the utility / accuracy of trained models. I used RMSE as a measure to identify and to choose good prediction models for further analysis.

Chapter 10: Data Evaluation

By using Auto Arima, VAR, Random forest, XG Boost, Multivariate deep learning model and LSTM for multivariate models, I did evaluation of right models using its RMSE results.

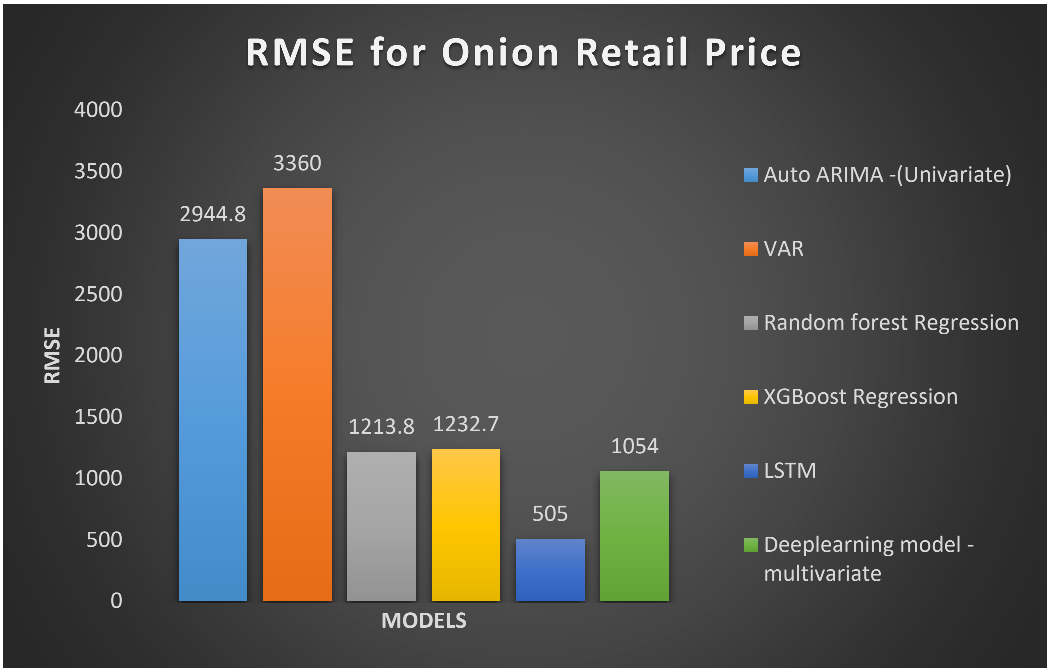

Figure 10.1: RMSE evaluation for onion retail price

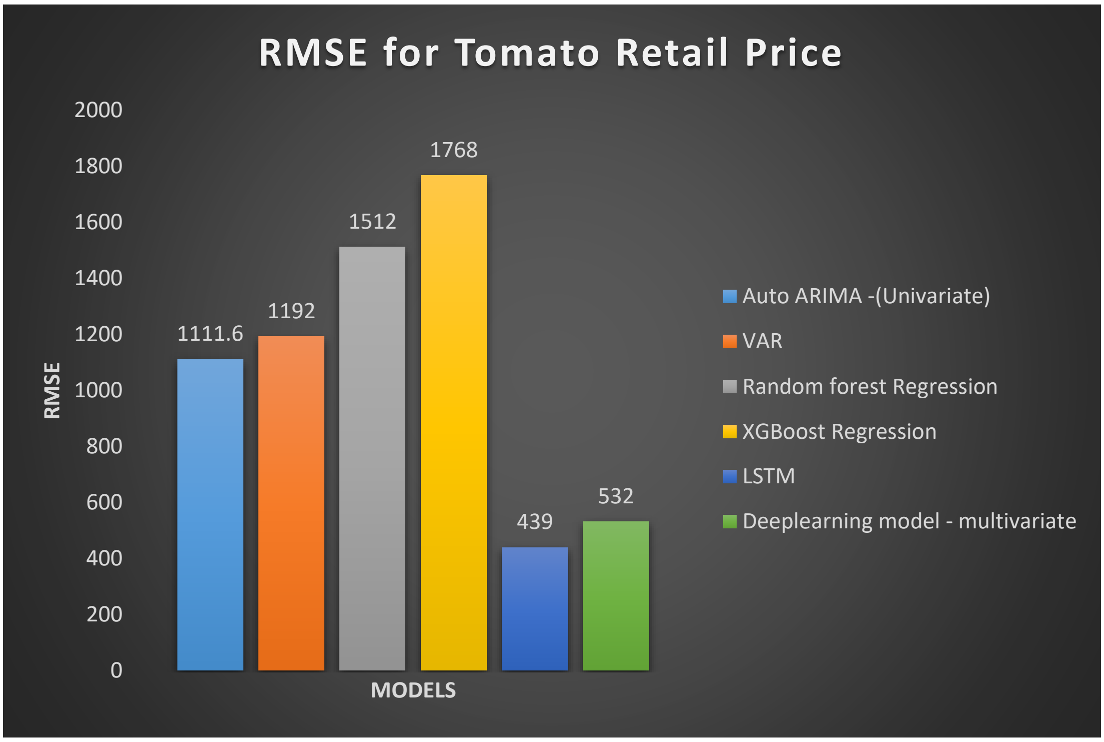

Figure 10.2: RMSE evaluation for tomato retail price

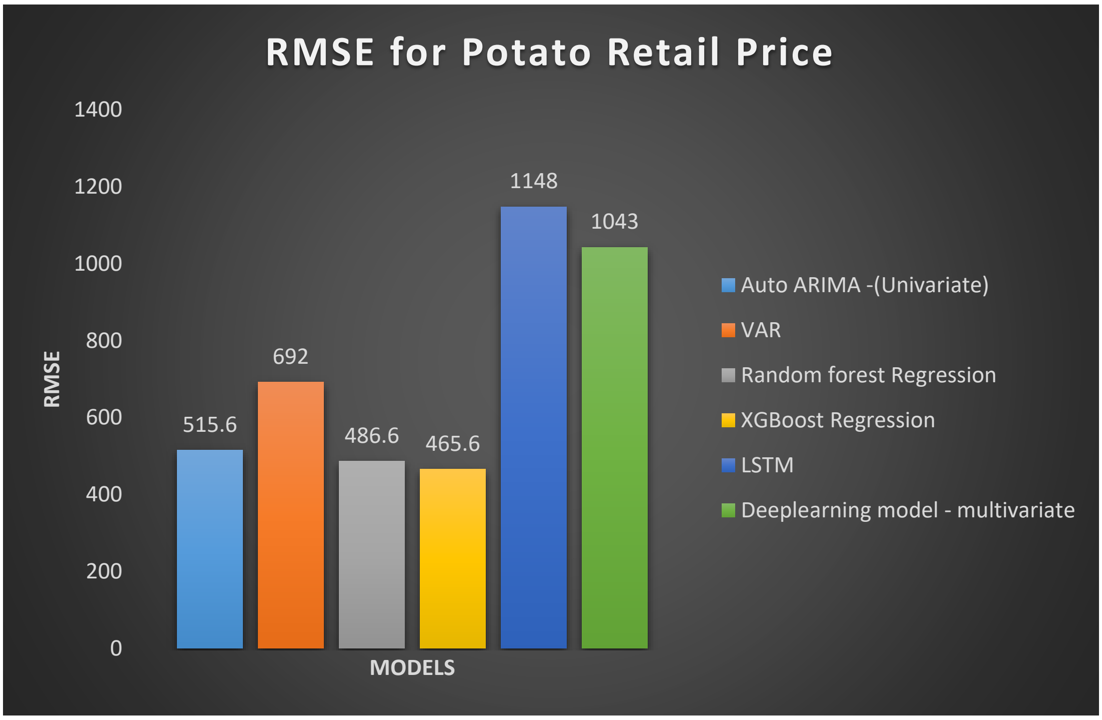

Figure 10.3: RMSE evaluation for potato retail price

Best 3 Models for onion, tomato and potato retail price predictions are as follows:

Onion Retail Price prediction: LSTM, Deep learning and Random forest

Tomato Retail Price prediction: LSTM, Deep learning and Random forest

Potato Retail Price prediction: XG Boost, Random forest and Auto ARIMA (Univariate).

Chapter 11: Deployment

Finally, we built different models and tested all the different metrics and are now ready to incorporate model in output.

Proposal for the deployment of models for the prediction of onion, tomato and potato are as follows:

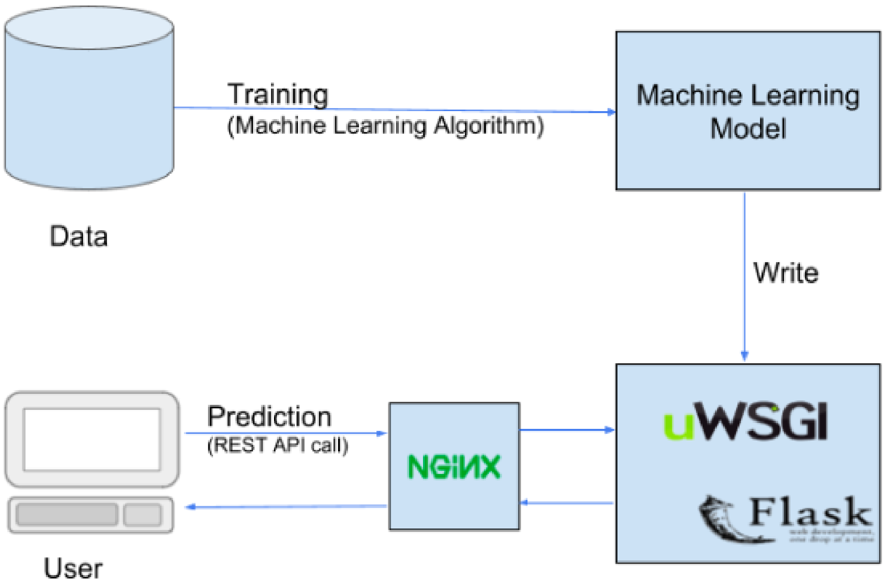

Figure 11.1: Data pipe line deployment architecture proposal (Gupta, 2019)

Deployment activities from this current project will carry it forward as a continuous activity in the next project.

Chapter 12: Analysis and Results

In this project we identified key models to predict multivariate prediction considering volume in metric tonnes and diesel price in rupees for onion, tomato and potato from January 2015 till July 2020.

During data understanding also observed linear relationship between wholesale and retail price of the onion, tomato and potato. I took a decision to drop whole sale price variable from study to avoid overfitting.

There are few major anomalies identified from the data for onion, tomato and potato and identified due to key events are as follows:

Onion Anomalies identified from the data during this project which happened during November 2019 and December 2019.

"Heavy unseasonal rainfall has locked latest onion produce in wet fields throughout onion growing states of Maharashtra, Karnataka and Telangana. As a ripple effect, onion prices increased everywhere in the country. The politically sensitive bulb's prices however are set to ease in coming months, sparking fears farmers will not get fair prices for their crop" (Times, 2019)

Tomato Anomalies identified from the data during this project which happened during July 2017, August 2017 and November 2017.

"Between July 2017 to November 2017 there was huge reduction in arrival of volume in different mandi's in India" (Hindu, 2019)

Potato Anomalies identified from the data during this project which happened during Nov 2018 and July 2020.

"Potato price increased continuously from July till December 2018" (Express, 2019)

"Supply shortage, rise in demand push potato prices in India up by 40% in major cities" (News, 2019)

Chapter 13: Conclusions and Recommendations for future work

Agricultural commodity onion, tomato and potato sales flows through the mandis in India.

Our study in this project help to understand the geological location of different mandi's based on overall volume of onion, tomato and potato from January 2020 till July 2020.

In this project we also understood the relationship of retail price with wholesale price, volume and diesel price of onion, tomato and potato.

Based on different statistical tests and data understanding we considered necessary variables for further prediction of Retail price which is the key output we are predicting in this project.

The prediction models selected also able to understand the data and pattern. we identified different models like Auto ARIMA, VAR, LSTM, Deep learning model, Random forest and XG Boost.

We identified Multivariate models is performing better than conventional univariate times series models.

We are also having potential opportunity to continue this work to further do deployment, rest API, web apps and also to include other variable which is affecting Retail prices of Onion, tomato and potato.

My key contribution in this Project is to emphasize on the role that of multivariate Machine learning prediction and other forecasting methods which will help in the future forecasting of agricultural commodities retail prices of onion, tomato and potato in India.

Bibliography

Agmarket. (2020). *agmarknet*. Agmarket. http://agmarknet.gov.in/PriceTrends/Default.aspx

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues*, *9*(5), 272–278.

Anil Seth. (2007). *Granger causality*. Scholarpedia. http://www.scholarpedia.org/article/Granger_causality#:~:text=Granger causality is a statistical,values of X2 alone.

Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, *7*(2), 1–21. https://doi.org/10.4018/ijoris.2016040101

Balaji Prabhu, B. V., Dakshayini, M., Istambul, M. R., Abinowi, E., Vohra, A., Pandey, N., Khatri, S. K., Balaji Prabhu, B. V., Dakshayini, M., Madaan, L., Sharma, A., Khandelwal, P., Goel, S., Singla, P., & Seth, A. (2018). Performance Analysis of the Regression and Time Series Predictive Models using Parallel Implementation for Agricultural Data. In *Procedia Computer Science* (Vol. 132, Issue 6). Elsevier B.V. https://doi.org/10.1016/j.procs.2018.05.187

Bhattacharya, A. (2020). *Anomalies*. Towardsdatascience. https://towardsdatascience.com/effective-approaches-for-time-series-anomaly-detection-9485b40077f1#:~:text=What is Time Series Anomaly,from rest of the data.

Bratsas, C., Koupidis, K., Salanova, J. M., Giannakopoulos, K., Kaloudis, A., & Aifadopoulou, G. (2020). A comparison of machine learning methods for the prediction of traffic speed in Urban places. *Sustainability (Switzerland)*, *12*(1), 1–15. https://doi.org/10.3390/SU12010142

Corporatefinanceinstitute. (2020). *Cointegration test*. Corporatefinanceinstitute. https://corporatefinanceinstitute.com/resources/knowledge/other/cointegration/

Express, F. (2019). *Anomaly 3*. Financial Express. https://www.financialexpress.com/economy/potato-prices-are-up-nearly-50-in-december-but-here-is-a-good-news/1442954/

Gati. (2020). *Diesel price*. Gati. https://www.gati.com/tools/view-fuel-diesel-surcharge/

Gupta, S. (2019). *Deployment*. Medium. https://medium.com/analytics-vidhya/deploying-a-machine-learning-model-on-web-using-flask-and-python-54b86c44e14a

Hindu. (2019). *Anomaly news2*. Hindu. https://www.thehindubusinessline.com/economy/agri-business/onion-tomato-price-spike-season-not-the-only-reason/article9957255.ece

India, H. department of. (2020). *Horticulture Data*. Nhb. http://nhb.gov.in/OnlineClient/MonthwiseAnnualPriceandArrivalReport.aspx

Istambul, M. R., Abinowi, E., Vohra, A., Pandey, N., Khatri, S. K., Balaji Prabhu, B. V., Dakshayini, M., Madaan, L., Sharma, A., Khandelwal, P., Goel, S., Singla, P., & Seth, A. (2019). Performance Analysis of the Regression and Time Series Predictive Models using Parallel Implementation for Agricultural Data. *Procedia Computer Science, 132*(6), 52–64. https://doi.org/10.1016/j.procs.2018.05.187

Li, X., Shang, W., Wang, S., Pandey, A., Gupta, V., Gupta, R. K., Madaan, L., Sharma, A., Khandelwal, P., Goel, S., Singla, P., Seth, A., Mango, N., Mapemba, L., Tchale, H., Makate, C., Dunjana, N., Lundy, M., Park, J. S., … Thampy, A. (2019). Editorial. In *IIMB Management Review* (Vol. 31, Issue 2). Elsevier Ltd. https://doi.org/10.1016/j.iimb.2019.09.001

Madaan, L., Sharma, A., Khandelwal, P., Goel, S., Singla, P., & Seth, A. (2019). Price forecasting & anomaly detection for agricultural commodities in India. In *COMPASS 2019 - Proceedings of the 2019 Conference on Computing and Sustainable Societies*. https://doi.org/10.1145/3314344.3332488

Murthy, B. (2019). *Onion Production*. The Print. https://theprint.in/opinion/how-india-can-ensure-onions-are-all-through-year-at-good-price/334477/

News, P. (2019). *Anomaly 4*. Potato News. https://www.potatonewstoday.com/2020/07/28/supply-shortage-rise-in-demand-push-potato-prices-in-india-up-by-40-in-major-cities/

Onion, C. O. F. (2016). *Supplementary Material : Price Forecasting & Anomaly Detection for Agricultural Commodities in India*.

Ouyang, H., Wei, X., & Wu, Q. (2019). Agricultural commodity futures prices prediction via long- and short-term time series network. *Journal of Applied Economics*, *22*(1), 468–483. https://doi.org/10.1080/15140326.2019.1668664

Shome, S., Khatri, U., Joshi, D., & Mehndiratta, S. (2018). Dynamic Fuel Pricing in India: An Event Study Methodology. *International Journal of Management Studies*, *V*(4(6)), 32. https://doi.org/10.18843/ijms/v5i4(6)/05

Subedar, S., Santosha, R., & Abhishek, S. (2017). Hybrid Time Series Modelling for Forecasting the Price of Washed Coffee (Arabica Plantation Coffee) in India. *International Journal of Agriculture Sciences Citation*, *9*(10), 4004–4007. http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217

Subhasree, M., Arun Priya, M. C., & Arun Priya, C. (2016). Forecasting Vegetable Price Using Time Series Data. In *International Journal of Advanced Research* (Vol. 3, Issue 535). http://www.journalijar.com

Times, E. (2019). *Anomaly news1*. Times. https://economictimes.indiatimes.com/news/economy/agriculture/flooding-in-major-onion-producing-states-has-led-to-spike-in-rices/articleshow/71975921.cms?from=mdr

Timothy Susanto. (2020). *XG Boost*. https://mc.ai/xgboost-time-series-for-forecasting-stocks-price/

Varun, R., Neema, N., Sahana, H. P., Sathvik, A., & Muddasir, M. (2019). Agriculture Commodity Price Forecasting using Ml Techniques. In *International Journal of Innovative Technology and Exploring Engineering* (Vol. 9, Issue 2S). https://doi.org/10.35940/ijitee.b1226.1292s19

wiki. (2020). *Augmented_Dickey test*. WIKIPEDIA. https://en.wikipedia.org/wiki/Augmented_Dickey–Fuller_test

WIKI. (2020). *RMSE*. WIKIPEDIA. https://en.wikipedia.org/wiki/Root-mean-square_deviation

Wikipedia. (2020a). *CRISP DM*. WIKIPEDIA. https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

Wikipedia. (2020b). *Onion In india*. WIKIPEDIA. https://en.wikipedia.org/wiki/Agriculture_in_India

Wikipedia. (2020c). *Potato*. Wikipedia. https://en.wikipedia.org/wiki/Potato

Wikipedia. (2020d). *Tomato*. WIKIPEDIA. https://en.wikipedia.org/wiki/Tomato

Wikipedia. (2020e). *VAR*. https://en.wikipedia.org/wiki/Vector_autoregression#:~:text=Vector autoregression (VAR) is a,type of stochastic process model.&text=This equation includes the variable’s,model%2C and an error term.

Zhang, D., Chen, S., Liwen, L., & Xia, Q. (2020). Forecasting Agricultural Commodity Prices Using Model Selection Framework with Time Series Features and Forecast Horizons. *IEEE Access*, *8*, 28197–28209.

https://doi.org/10.1109/ACCESS.2020.2971591

Appendix

Plagiarism Report[1]

"Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato"

by Suresha Hp

Submission date: 24-Oct-2020 04:19PM (UTC+0530)
Submission ID: 1425124193
File name: mparative_study_of_forecasting_methods_Suresha_HP_23Oct2020.docx (2.2M)
Word count: 6447
Character count: 37302

[1] Turnitin report to be attached from the University.

"Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato"

ORIGINALITY REPORT

6%	5%	1%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Southern Queensland Student Paper	1%
2	Submitted to Sogang University Student Paper	1%
3	Submitted to Siddaganga Institute of Technology Student Paper	1%
4	journalofbigdata.springeropen.com Internet Source	<1%
5	www.termpaperwarehouse.com Internet Source	<1%
6	Submitted to Lovely Professional University Student Paper	<1%
7	Submitted to Regional Centre for Biotechnology Student Paper	<1%
8	repositorio-aberto.up.pt Internet Source	<1%

9	www.slideshare.net Internet Source	<1%
10	www.inmybangalore.com Internet Source	<1%
11	Haoyang Wu, Huaili Wu, Minfeng Zhu, Weifeng Chen, Wei Chen. "A new method of large-scale short-term forecasting of agricultural commodity prices: illustrated by the case of agricultural markets in Beijing", Journal of Big Data, 2017 Publication	<1%
12	student-friendly.blogspot.com Internet Source	<1%
13	towardsdatascience.com Internet Source	<1%
14	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	<1%
15	Lovish Madaan, Ankur Sharma, Praneet Khandelwal, Shivank Goel, Parag Singla, Aaditeshwar Seth. "Price forecasting & anomaly detection for agricultural commodities in India", Proceedings of the Conference on Computing & Sustainable Societies - COMPASS 19, 2019 Publication	<1%
16	Submitted to uvt Student Paper	<1%

Publications in a Journal/Conference Presented/White Paper[2]

Selected for CST 2020, Zurich Switzerland

Dear Author,

First of all, thank you very much for submitting your paper to **CST 2020** to be held in **Zurich, Switzerland**, **November 21 ~ 22, 2020**. Based upon the reviewer's reports, we are pleased to inform you that your paper has been **ACCEPTED** by the conference and will be included in the proceedings published by Computer Science Conference Proceedings in Computer Science & Information Technology (CS & IT) series.

Congratulations on your excellent work!

In order to achieve the highest quality proceedings, we urge you to carefully consider the reviewer's comments, if any, when preparing the final version of your paper.

1. Please read the following Information carefully to prepare a final manuscript of your paper https://cst2020.org/submission/template.doc.

The maximum number of pages without extra payment is 20 (CCSP format). For each extra page you have pay 50 USD additionally.

2. Submit your final camera-ready version of paper (.doc version+ .pdf version) and filled CR form **cst@cst2020.org (or) cst.conf@yahoo.com**

3. Final Manuscript Submission Details:

a) When submitting your final manuscript, please ensure that you send us all source files such as .doc and pdf.

Any Additional Details

Due to high fees not participated in CST 2020. Submitted 2nd International Conference on Advances in Distributed Computing and Machine Learning (ICADCML-2021).

[2] URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided.