BAN 5600 - Advance Big data Computing and Programming

Final Project

Price trend prediction for world food prices for low-middle income countries

By

Shashank Gupta

Shivika Gupta

Reeha Naaz Hindustan

Project link: Github

1.0 Abstract

This report analyses the effect of the food price trend in low to middle income countries covering Asia, Africa and some countries from South America for the period 2003 to 2020. The study uses data obtained from the World Food Programme of United Nations' to achieve Sustainable Development Goals. The data is used to forecast price trends for different commodities in the given countries for the next 2-3 years, evaluate suitable modes to predict the food price trends as well as use mapping tools to study how market size has changed over time and so on.

2.0 Introduction

The lifecycle for the project to fight hunger typically involves Country strategic planning, analyses and assessments, programme design, types of support, supply chain and monitoring, evaluation & learning. Food price trend data is analyzed to support the above milestones.

Food prices refer to the average price of particular food commodities globally and across countries. Commodities usually include rice, wheat, sugar, maize etc. and the demand varies for each country. The price of goods not only provides an important indicator of the balance between agricultural production and market demand, but also have strong impacts on food affordability and income. Food prices not only influence consumer affordability, but also influence the income of farmers and producers. In low-to-middle income countries in particular, a large share of the population is employed in agriculture. Producers typically benefit from higher food prices and consumers from lower prices. Food markets can therefore have a strong impact on food affordability, hunger and undernourishment and dietary quality.

How swiftly the world market for food can change could be observed in the mid-2000s. For two decades, leading up to the millennium, global demand for food had increased steadily, along with growth in the world's population, record harvests, new technologies, improvements in incomes, and the diversification of diets. Food prices continued to decline through 2000. However, in 2004, prices for most grains began to rise. Rising production could not keep pace with the even stronger growth in demand. Food stocks became depleted. And then, in 2005, food supply was squeezed by disappointing harvests in major food-producing countries. By 2006, world cereal production had fallen by 2.1 per cent. In 2007, rapid increases in oil prices increased fertilizer and other food production costs

3.0 Literature Review

The World Food Programme (WFP), a United Nations subsidiary, is the leading humanitarian organization saving lives and changing lives, delivering food assistance in emergencies and working with communities to improve nutrition and build resilience. As the international community has committed to end hunger, achieve food security and improved nutrition by 2030, one in nine people worldwide still do not have enough to eat. Food and food-related assistance lie at the heart of the struggle to break the cycle of hunger and poverty. For its efforts to combat hunger, for its contribution to bettering conditions for peace in conflict-affected areas and for acting as a driving force in efforts to prevent the use of hunger as a weapon of war and conflict, WFP was awarded the Nobel Peace Prize in 2020. In 2019, WFP assisted 97 million people – the largest number since 2012 – in 88 countries. On any given day, WFP has 5,600 trucks, 30 ships and nearly 100 planes on the move, delivering food and other assistance to those in most need. Every year, more than 15 billion rations at an estimated average cost per ration of US$ 0.61 is distributed by WFP. These numbers lie at the roots of WFP’s unparalleled reputation as an emergency responder, one that gets the job done quickly at scale in the most difficult environments.

The dataset used for this project contains Food Prices data for all countries. The data comes from the World Food Programme (WFP) and covers foods such as maize, rice, beans, fish, and sugar for 76 countries and some 1,500 markets. It is updated weekly but contains to a large extent monthly data. The data goes back as far as 1990 for a few countries, although many countries started reporting from 2003 or thereafter.

4.0 Methodology

CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is an industry-proven way to guide data mining efforts. By following the methodology of the CRISP-DM process, every step of the project is defined. The following diagram explains it more:

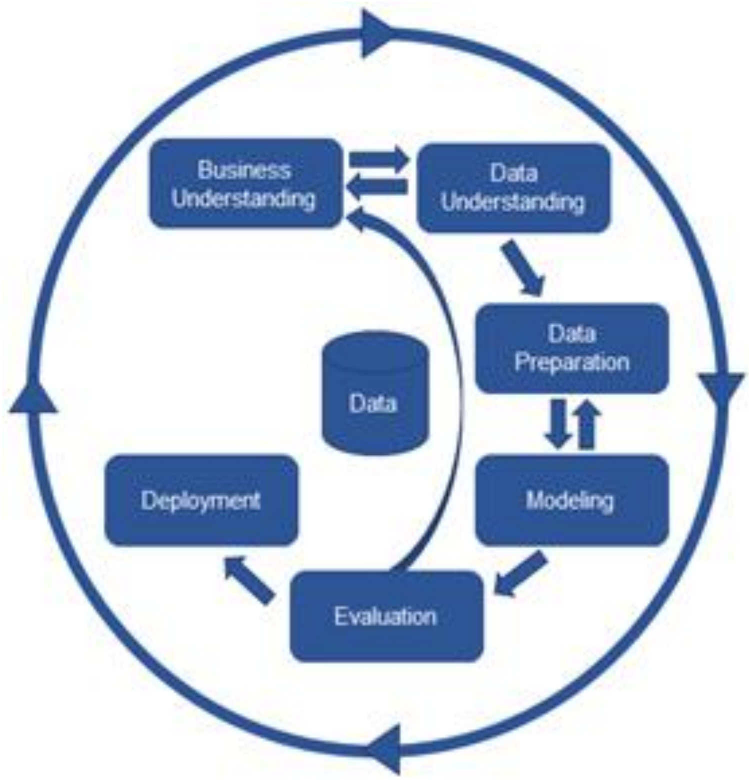

Fig-1

To explain the process further, Business Understanding involves the study of the business domain. In this case, we study the Food price data for all the countries that come from the World Food program. Data Understanding deals with the study of the various independent variables in the dataset and the dependent variable. Data cleaning is followed by data understanding. Here, we deal with the missing values, outliers, and NA values because we want to understand each of the variables, the correlations among them, and also their relationship with the predictor variable, we proceed with Exploratory Data Analysis.

4.1 About Exploratory Analysis

This also marks the final step in which we prepare the data for the modeling step. Modeling is the step wherein the technique used for modeling is identified and stated along with any assumptions if required. In this case, Linear regression, Decision Tree Regression, and Random Forest Regression algorithms are identified. This is also the stage where the dataset is divided into testing and training datasets. The model is built and assessed on the training dataset first. The parameters are well defined and revised according to the need after the assessment procedure.

Evaluation is the stage where data mining results are assessed, and approved models are finalized for further implementation. This is the stage where the entire process is summarized and those activities which were missed are highlighted for use in the future. All those activities that are approved to be used further are also stated to finish finalizing the exact modeling procedure. The decision on how to proceed is stated along with the rationale.

Deployment is the final step in the process wherein we summarize the strategy to deploy the project and how to perform the steps to implement in the real-world.

4.2 Link to the dataset

This dataset contains 480,239 rows and 13 attributes such as country, market, commodity, year, month, alert of price spikes, unit of measure, price type and currency which are the dependent variables, and price trend being the independent variable (target variable). Main aim of this project is to form structures and patterns to draw inferences about the food price trend across the middle-income countries and analyze how price trends have influenced the markets and the availability of commodities in each country over the years.

4.3 Data Description

Variable	Description
Country	Country name
mkt_id	Market identification number
Market	Market name
CommodityGroup	Commodity Group
Commodity	Food item (rice, wheat, etc.) in the commodity group
Year	Year of price trend
PriceTrend	Price trend of the commodity (Target Variable)
Month	Month of price trend
PEWI	Unknown index
ALPS	Alert of Price Spikes (Alert, Crisis, Stress, Normal)
UnitofMeasure	Unit of measure (Kilogram, liter)
PriceType	Price type (Retail/Wholesale)
Currency	Currency of the country variable
sn	Serial number

Table 1

4.4 Exploratory Data Analysis

4.4.1 Average Price trend across all the countries from 2003-2020

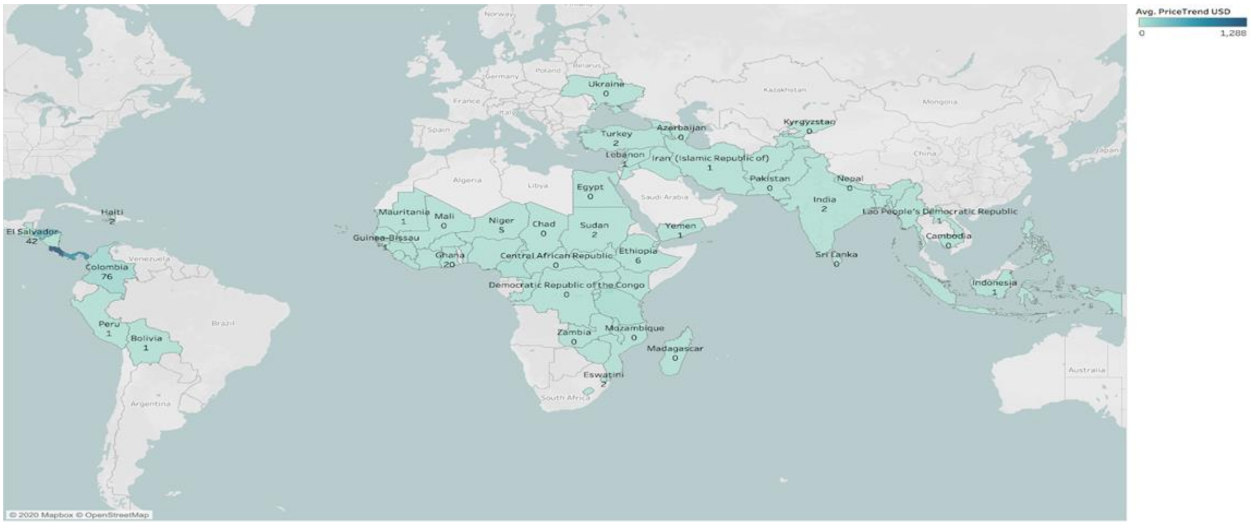

Fig 2

Price trends were mapped for all the low-income and middle-income countries to understand the average price trend for each country over the years. In South America, Columbia and El Salvador have high average prices compared to all other countries in the world. In Africa, Ghana has the highest average price of $20/kg followed by Ethiopia and Niger with $6 and $5 per kg. In Asia, most countries have an average price ranging from $0-2 per kg.

4.4.2 Market count in each country

Fig 3

Average price trends can vary based on many factors and one such factor is the market size. In South America, the highest number of markets are observed in Columbia, Haiti and Bolivia, all having 9 markets each. In Africa, market size is quite large for countries like Niger having 68, Ethiopia 56, Chad 33 and Madagascar 31. In Asia, India and Sri Lanka are leading with 60 and 25 markets each. Factors such as population density, supply-demand, agricultural factors for crop growth, etc. affect the overall market size.

4.4.3 Volume of Commodity groups in each country

Fig 4

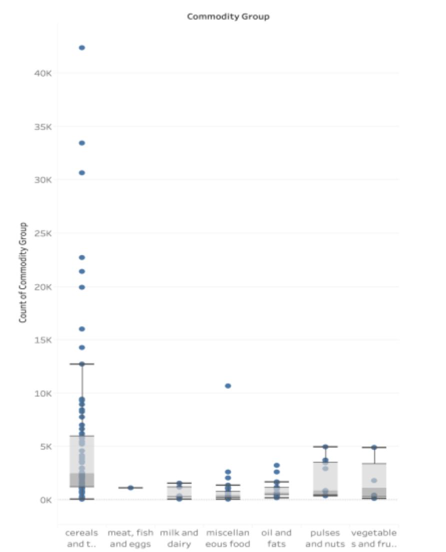

Fig 5

Total of 7 Commodity groups are analyzed in this plot. They are - 1. cereals and tubers 2. meat, fish and eggs 3. Milk and dairy 4. Oil and fats 5. Pulses and nuts 6. Vegetable and fruits 7. Miscellaneous food. It is observed that more than 60% of the commodities sold come under cereals and tubers across all the countries.

4.4.4 Change in Market size based on commodity group over the years

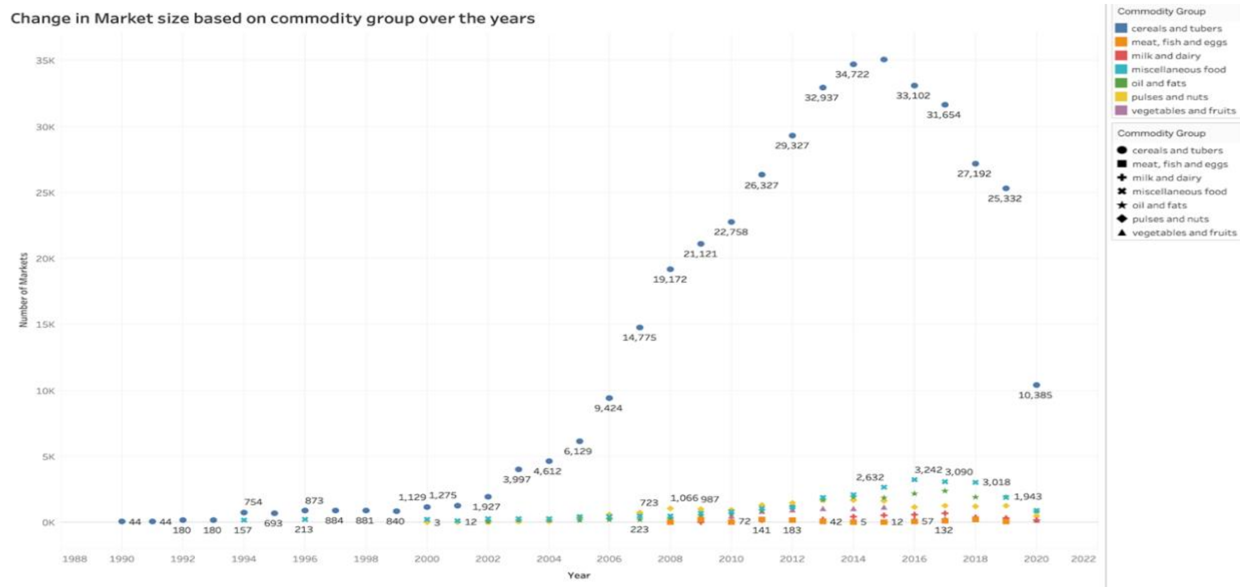

Fig 6

It is observed that cereals and tubers have always been on demand across most of the markets around the world. Primarily because cereals and tubers contain essential crops for consumption. Some of those include rice, sorghum, sweet potatoes, teff, tortilla, wheat and yam. Most countries are also soil friendly to these crops due to which, availability of cereals and tubers has increased over the years.

4.4.5 Forecast of Price trend for Niger based on the commodities sold.

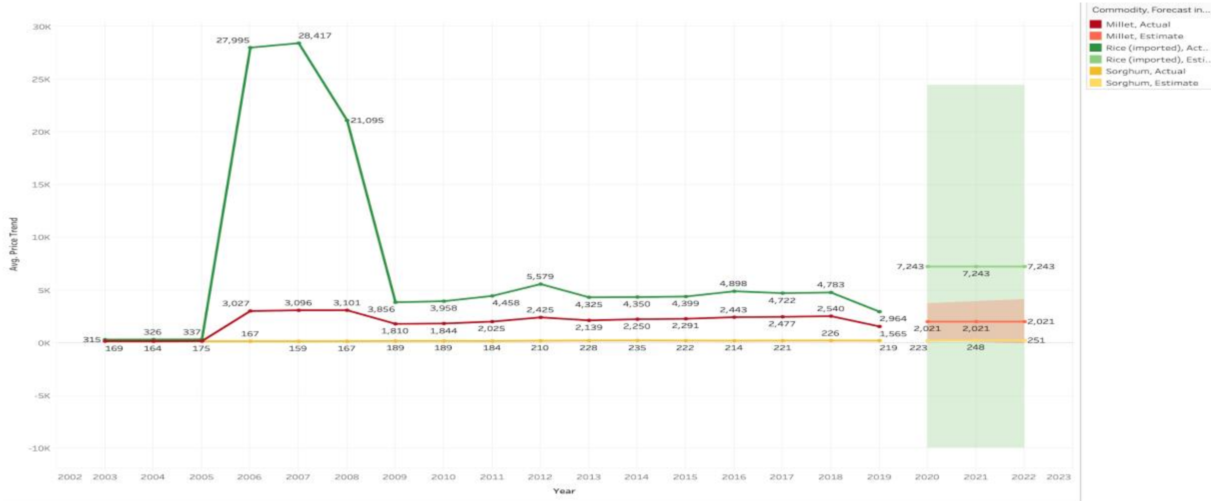

Fig 7

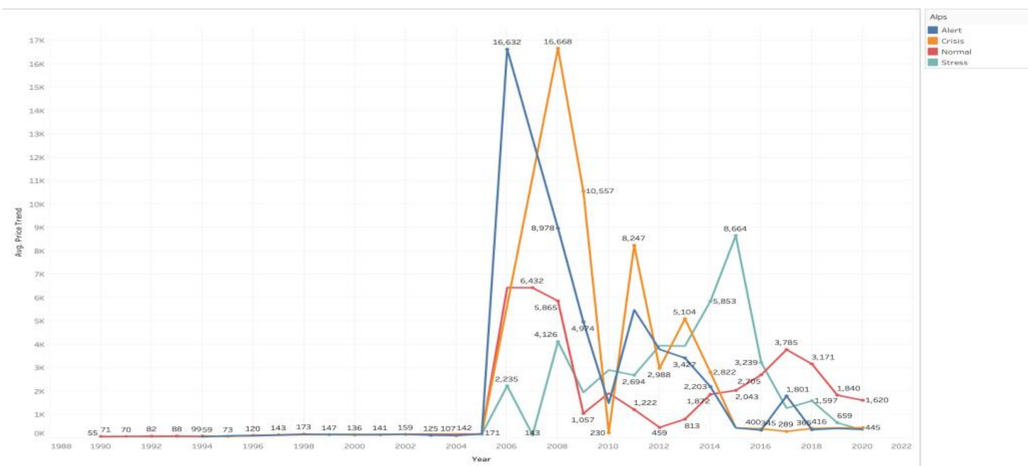

Fig 8

Niger has the largest market size amongst all the countries and is hence used to forecast price trends for 2020-2022 for the type of commodities sold for the group cereals and tubers. Rice is the most consumed food type and is predicted to have higher prices compared to the other commodities such as millet and sorghum. This high range for rice is probably observed due to ‘Alert’ in price spikes for the market in 2006 and ‘Crisis’ in 2007 and 2008 for most of the months. The market position (ALPS) is usually categorized based on market price and estimated price over a certain time period.

4.4.6 Alert of Price Spikes (ALPS) forecast for Niger

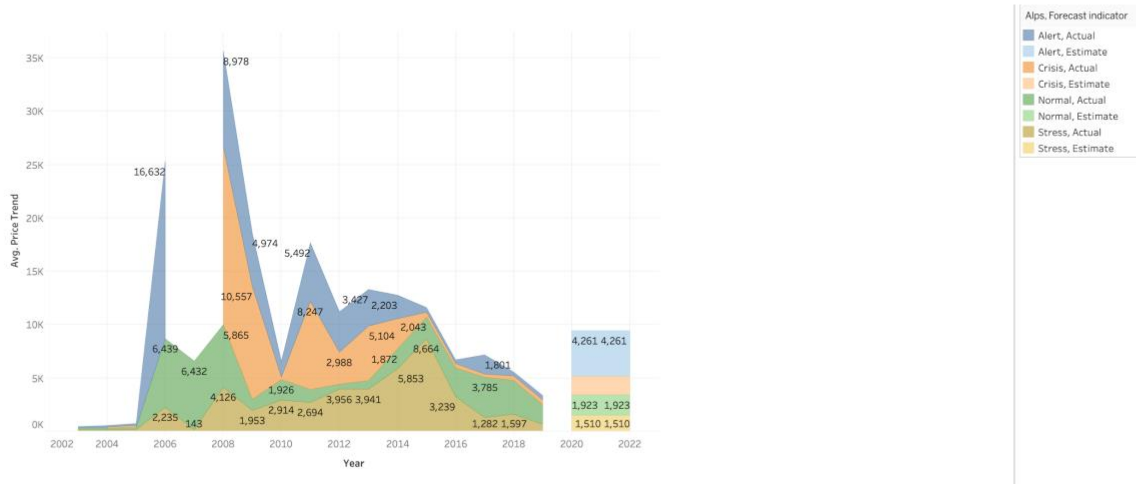

Fig 9

Based on the past data for ALPS, forecast predicts lower prices for Crisis, Normal and Stress position of the markets in Niger compared to past years. In the state of Alert, Price trend is relatively higher for the years 2020-2022. This can be further analyzed using

<u>ALPS is calculated by the following formula -</u>

ALPS = (Observed market price - Estimated price) / Standard deviation of residuals

Situation on a given market :	**ALPS thresholds :**	**Interpretation in terms of price gap :**
Normal	ALPS < 0.25	$(Price_{it} - \widehat{Price_{it}})/\ \sigma_{\hat{\varepsilon}} < 0.25$
Stress	0.25 ≤ ALPS < 1	$0.25 \leq (Price_{it} - \widehat{Price_{it}})\ /\sigma_{\hat{\varepsilon}} < 1$
Alert	1 ≤ ALPS < 2	$1 \leq (Price_{it} - \widehat{Price_{it}})\ /\sigma_{\hat{\varepsilon}} < 2$
Crisis	ALPS ≥ 2	$(Price_{it} - \widehat{Price_{it}})/\sigma_{\hat{\varepsilon}} \geq 2$

Fig 10

5.0 Results

After completing the analysis, machine learning algorithms such as linear regression, decision tree regression and random forest regression were implemented to perform price predictions. Each model was evaluated using RMSE score to select the model most appropriate to perform food price trend prediction.

The training and testing dataset is split into the ratio 7:3, i.e., 70% of the data is used for training and 30% is used for testing.

Linear regression

Normalized Root Mean Squared Error (NRSME) = RMSE / Mean (Price_trend_USD)

= 37.165 / 5.48

= 6.782

Country	market	Year	CommodityGroup	features	Price_trend_USD	prediction
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	38.57960214313141	44.373795412187064
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	37.895209119838846	44.35845148756016
India	Chennai	2020	cereals and tubers	[931.0,2020.0,0.4...	39.91396338931951	44.3396719673849
India	Chennai	2020	cereals and tubers	[931.0,2020.0,2.3...	40.431054985502776	44.29305796530889
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.353422010932574	38.05927476352008
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.578968985592526	38.026637308575346
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	10.111704163047854	37.987438485922894
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	13.461085566582264	37.94802167444004
Mali	Bamako	2006	cereals and tubers	[247.0,2006.0,-0....	0.31531039034063996	-1.5260708777342415
Mali	Bamako	2011	cereals and tubers	[247.0,2011.0,0.9...	0.39093259874699826	-0.08947695354834195
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.3...	0.43036759990498064	0.5052255292380323
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.8...	0.3985157415625715	0.3571291534484544
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,-0....	0.2354308992723246	-2.742159744599917
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.0...	0.2957227417014005	-2.7864392911986897
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.2565651678695927	-2.8170251080192656
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.29512233225252904	-2.818888280460669

Fig 11

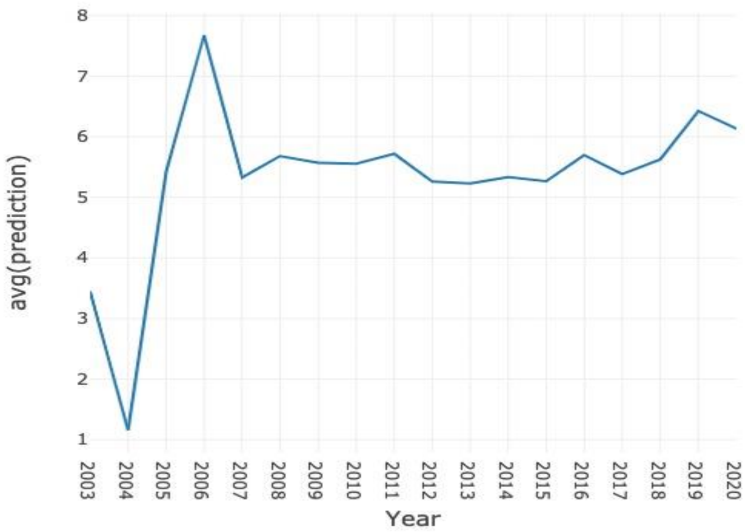

Fig 12

Decision Tree Regression

NRMSE = RMSE / Mean (Price_trend_USD)

= 24.1235 / 5.48

= 4.402

Country	market	Year	CommodityGroup	features	Price_trend_USD	prediction
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	38.57960214313141	47.95304117914907
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	37.895209119838846	47.95304117914907
India	Chennai	2020	cereals and tubers	[931.0,2020.0,0.4...	39.91396338931951	47.95304117914907
India	Chennai	2020	cereals and tubers	[931.0,2020.0,2.3...	40.431054985502776	47.95304117914907
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.353422010932574	19.782595783730546
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.578968985592526	19.782595783730546
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	10.111704163047854	19.782595783730546
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	13.461085566582264	19.782595783730546
Mali	Bamako	2006	cereals and tubers	[247.0,2006.0,-0....	0.31531039034063996	0.7574027078654015
Mali	Bamako	2011	cereals and tubers	[247.0,2011.0,0.9...	0.39093259874699826	3.816779676250239
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.3...	0.43036759990498064	3.816779676250239
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.8...	0.3985157415625715	3.816779676250239
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,-0....	0.2354308992723246	0.7574027078654015
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.0...	0.2957227417014005	0.7574027078654015
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.2565651678695927	0.7574027078654015
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.29512233225252904	0.7574027078654015

Fig 13

Fig 14

Random Forest Regression

NRMSE = RMSE / Mean (Price_trend_USD)

= 26.2952 / 5.48

= 4.798

Country	market	Year	CommodityGroup	features	Price_trend_USD	prediction
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	38.57960214313141	42.618204505502256
India	Chennai	2019	cereals and tubers	[931.0,2019.0,-0....	37.895209119838846	42.618204505502256
India	Chennai	2020	cereals and tubers	[931.0,2020.0,0.4...	39.91396338931951	42.10707161512664
India	Chennai	2020	cereals and tubers	[931.0,2020.0,2.3...	40.431054985502776	39.03874151944733
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.353422010932574	24.043035492903485
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	14.578968985592526	24.523456057604548
India	Delhi	2003	cereals and tubers	[934.0,2003.0,-0....	10.111704163047854	23.96894996893146
Mali	Bamako	2006	cereals and tubers	[247.0,2006.0,-0....	0.31531039034063996	2.0603347512042425
Mali	Bamako	2011	cereals and tubers	[247.0,2011.0,0.9...	0.39093259874699826	2.512931585969737
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.3...	0.43036759990498064	2.5224079970409417
Mali	Bamako	2012	cereals and tubers	[247.0,2012.0,0.8...	0.3985157415625715	2.512931585969737
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,-0....	0.2354308992723246	1.9669043478800998
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.0...	0.2957227417014005	1.9669043478800998
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.2565651678695927	1.9669043478800998
Mali	Gao	2003	cereals and tubers	[248.0,2003.0,0.1...	0.29512233225252904	1.9669043478800998

Fig 15

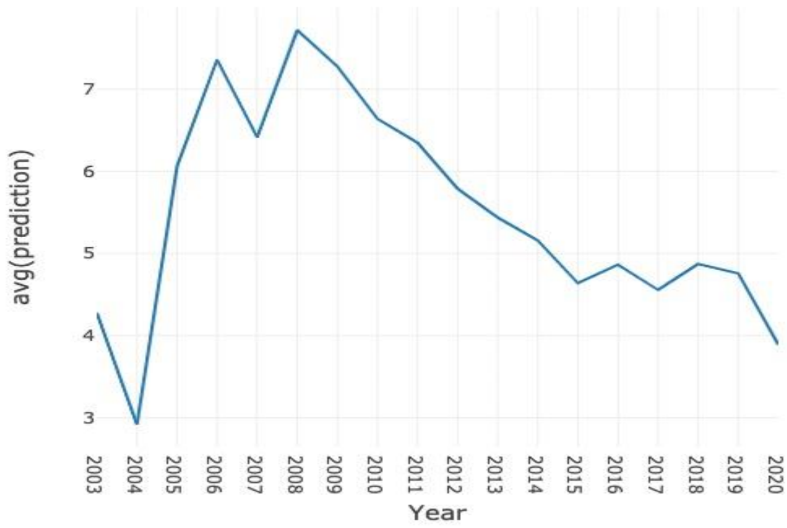

Fig 16

Summary

Model	**Normalized root mean squared error**
Linear Regression	6.782
Decision Tree Regression	4.402
Random Forest Regression	4.798

Table 2

6.0 Conclusion

- Root mean squared error (RMSE) is used to calculate the fitness of a regression model. The closer the value is to zero the better it would perform predictions.

Normalized RMSE (NRMSE) has been calculated to standardize the RMSE and compare it for different regression models.

- It is observed that the Decision tree model has the lowest normalized RMSE i.e. 4.402 as compared to the other models hence, this model fits the data best as compared to Linear regression and Random Forest regression models.
- We can conclude based on the predictions and exploratory data analysis, there will be an upward trend of price spike for different commodities.

7.0 Limitations and future scope

To study the effect of price trends in depth, certain information such as total population, population density and supply-demand information could be obtained from world bank data. This would help further analyze consumption of commodities through various markets over time. Along with these variables, the ALPS data available in our dataset can be used to predict the market position by running classification models. Detailed Time-series forecasting could be performed using models such as ARIMA, SVM, etc. for better predictions.

8.0 Business Recommendations

1. Based on the data for ALPS, the countries can be prepared for the types of alerts notified. While a low-price trend is observed for the 2020-2023 forecast for alert, crisis, normal and stress, the government can use this data to allocate budget for different markets based on the commodity group sold.
2. Demand for cereals and tubers has always been high and hence, markets all over the low to middle income countries must make sure they are able to provide enough of this commodity group.
3. Based on the agricultural factors, more amount can be invested by the government on commodities whose forecasted price is low compared to other commodity types. For instance, in Niger, Millets and Sorghum prices are almost

the same whereas price for rice is quite high. Based on the demand and budget, more can be spent on millets and sorghum and slightly lesser on rice for Niger.

4. Based on the predictions, most of the markets will be expecting a price increase for different Commodities and Countries and their markets should be warned by the UN about this change to cope up with this and try to eradicate hunger.

9.0 References

1. https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp264186.pdf
2. https://data.humdata.org/dataset/wfp-food-prices-for-the-world
3. https://dataviz.vam.wfp.org/economic_explorer/reports?adm0=181
4. https://www.wfp.org/
5. https://docs.aws.amazon.com/machine-learning/latest/dg/tutorial.html
6. https://docs.microsoft.com/en-us/azure/machine-learning/