Forecasting Commodity Prices

1st G M VINAY SREEKAR REDDY
CSE Bachelor of Technology
Presidency University
Bangalore, India
sreekarreddy1436@gmail.com

2nd SAHANA GOWDA KR
CSE Bachelor of Technology
Presidency University
Bangalore, India
krsahanagowda@gamil.com

3rd BHAVANA V
CSE Bachelor of Technology
Presidency University
Bangalore, India
vbhavana37@gmail.com

4th ANKITHA MORE B
CSE Bachelor of Technology
Presidency University
Bangalore, India
Anikithamore13@gmail.com

5th POOJITHA G
CSE Bachelor of Technology
Presidency University
Bangalore, India
poojithag842@gmail.com

Abstract

A thorough project report on the study and forecasting of commodities prices via the application of machine learning and linear regression methods is summarized in this abstract. The paper explores the importance of commodity price forecasting for agricultural decision-making, discussing how changes in pricing affect everyday life and agricultural productivity. The significance of comprehending the world's commodities markets is examined, with a focus on the use of machine learning algorithms specifically, linear regression for time series forecasting. To forecast correlations between variables, the process includes procedures like feature extraction, selection, and classification. The predicted results, such as R-squared values, trend detection, and risk assessment, are presented along with a suggested system that uses machine learning to classify and comprehend data from the commodities market. The usefulness of a thorough model selection framework is discussed in the report's conclusion, with an emphasis on the principle components technique and linear regression for commodity price prediction.(Food and Agriculture Organization of the United Nations 2013)

Keywords— Commodity prices, Agricultural authorities, Commodity market data, forecasting, linear regression, Python, Jupyter Notebook.

Introduction

The fluctuations in agricultural commodity prices have a profound impact on the inputs and outputs of agricultural production, as well as on people's daily lives. Consequently, a reliable forecast of commodity prices is essential for agricultural authorities to make educated decisions. The ability to foretell costs is required. (Food and Agriculture Organization of the United Nations 2021)When agricultural commodity prices rise, the quality of life of the poor falls, and vice versa when commodity prices fall, especially in developing countries. On the other hand, as prices rise or fall, those who work on making these things will feel the effects. Using a computational general equilibrium approach, this study investigates the effects of high and low commodity prices on agricultural output. This work introduces a novel method for selecting models with flexible consideration of time series features and prediction boundaries. We demonstrate that, for low-income persons, using data from a variety of typical commodity markets (wheat, paddy, common, rice, maize, jawar, sorghum, etc.) between 2022 and 2023(International Labour Office et al. 2004)

Problem Statement

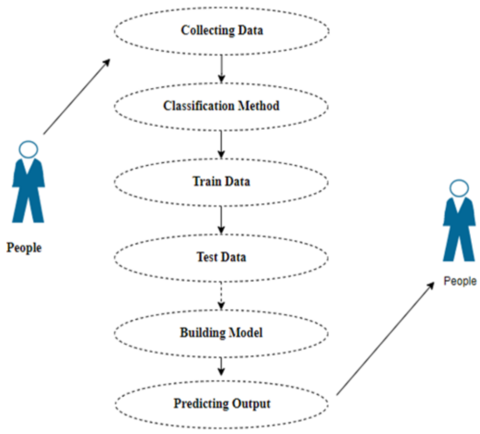
- The volatility in agricultural commodity prices poses significant challenges for agricultural authorities in making informed decisions. The absence of reliable forecasting models that effectively analyze and predict these price fluctuations hampers the ability to mitigate risks and optimize agricultural output. Existing methodologies often lack comprehensive approaches in utilizing machine learning, specifically linear regression, for accurate time series forecasting of commodity prices across diverse markets. Moreover, the absence of a unified model selection framework hinders the development of robust prediction models that cater to the complexities of global commodity markets. Consequently, there is a pressing need to develop an integrated and efficient system that leverages machine learning techniques to create reliable models capable of forecasting agricultural commodity prices accurately, thereby aiding decision-makers in risk assessment, trend identification, and optimizing agricultural outputs.

EXISTING METHOD:

- While most authors use US dollars and actual US interest rates for measuring commodity prices, we stick to the literature and use US dollars instead. Regarding commodity pricing in other countries, we believe this is a fair course of action as well.
- If transportation and trade restrictions divide commodity markets on a national level, then domestic factors like real interest rates and economic activity set commodity prices on a local level.
- But, it's safe to presume that global commodities markets are more linked than separated. It is often believed that commodities are subject to the law of one price. (Financial Crisis Inquiry Commission 2011) To convert US dollars to Australian dollars, we take the US dollar price of wheat and multiply it by the nominal exchange rate. The real cost of wheat in Australia is calculated by multiplying the actual cost in the US by the real rate of exchange.
- We take a different tack when it comes to India, as the nation produces a disproportionately large amount of white(International Labour Office et al. 2004) goods compared to other countries. Our rates are the most competitive when compared to those in other nations. Owing to the substantial demand for Chinese products, the cost of raw food items will be comparatively cheaper in China than in other nations.
- To prevent errors or outliers, it is essential that we adhere to the proper procedures. To avoid making mistakes, we must thoroughly research the agricultural practices of each nation and take appropriate measures accordingly.
- This problem statement highlights the challenges faced in forecasting agricultural commodity prices, including the absence of robust forecasting models, limitations in existing methodologies, and the necessity for a unified model selection framework to address the complexities of global commodity markets. (Food and Agriculture Organization 2019)

Methodology:

Machine learning has been used for algorithm recommendation jobs for a considerable duration. Additionally, since 2004, it has been explored in the field of time series forecasting. In this particular instance of Machine Learning, the focal point is the correlation between the characteristics of the data and the effectiveness of the algorithm. Regression is often used to get knowledge on such correlation. An experimental methodology is proposed for model selection in agricultural commodity price (Chang, Hsu, and Chang 2019)time series forecasting, using Machine learning. This study involves three primary steps: feature extraction, feature selection, and classification. The data was used to educate and enhance machine learning systems. The data was used for model training, while the remaining twenty percent was utilized for testing. In this investigation, we use linear regression. Regression analysis is a statistical method used to quantify the level of correlation between two or more variables. Regression analysis enables the investigation of how a change in one independent variable affects the value of the dependent variable while holding all other variables constant. (Brownlee 2017)



Proposed method:

- To overcome the obstacles given by the pervasiveness of financial data, the suggested approach seeks to create a machine learning algorithm that can categorize commodities data using statistical analysis. Processing and understanding this massive information manually have been more of a challenge as financial data becomes more accessible on more platforms.

- Machine learning provides a realistic way to automate the ordinarily laborious task of analyzing and classifying massive amounts of commodity data. The goal is to use machine learning to sort data on agricultural commodities into distinct groups so that market trends and patterns may be better understood. The model may learn statistical patterns and linkages between numerous market indicators by training it on historical data of agricultural commodities. This allows it to improve its accuracy over time.(Cartwright 2020)

- This method not only simplifies analysis but also makes it possible to adjust the model based on changing market conditions and trends. To choose the best method for agricultural commodities data classification, the machine learning model will test out several algorithms, tweak parameters, and compare their results. The model is continuously optimized and improved via this iterative process, guaranteeing its trustworthiness in capturing varied market trends and patterns.(OECD et al. 2010)

- We want to install a strong machine learning model that can automatically classify data from agricultural commodities. This model will be a great tool for tracking and analyzing market movements. To better understand market dynamics and make informed decisions, this method provides a scalable and efficient way to analyze massive amounts of financial data in real-time.(Pearson 2017)

Algorithm:

- It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn.
- It can use this test harness as a template for your machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods

like cross-validation, you can get an estimate of how accurate each model may be on unseen data.

- It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques to look at the data from different perspectives.
- The same idea applies to model selection. You should use several different ways of looking at the estimated accuracy of your machine learning algorithms to choose the one or two to finalize.
- A way to do this is to use different visualization methods to show the average accuracy, variance, and other properties of the distribution of model accuracies.
- In the next section you will discover exactly how you can do that in Python with scikit-learn. The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data and it can achieve this by forcing each algorithm to be evaluated on a consistent test harness.

Process Of Machine Learning

Linear Regression

- The statistical method known as linear regression is used to represent the relationship that exists between a dependent variable and a number of independent factors. The model operates on the assumption that the features of the input have a linear link with the variable that is being targeted.
- The objective of the model is to identify the straight line that provides the most accurate representation of the data; hence, it depicts this link as a straight line.

i. The variable that we are attempting to predict or explain is referred to as the dependent variable (Y).

ii. The variable that is utilized for making predictions is referred to as the independent variable (X).

iii. The assumption of a linear relationship is that changes in the independent variable (or variables) are related to a consistent change in the variable that is being studied (the dependent variable).

iv. The regression line is the line that most closely corresponds with the data points, hence minimizing the difference between the values that were predicted and those that were actually observed.

v. In order to establish the values that constitute the regression line, the model performs calculations to find the coefficients, namely the slope and the intercept.

vi. The mean squared error, often known as MSE, is a statistic that is used to assess the degree to which a model accurately predicts the variable of interest.

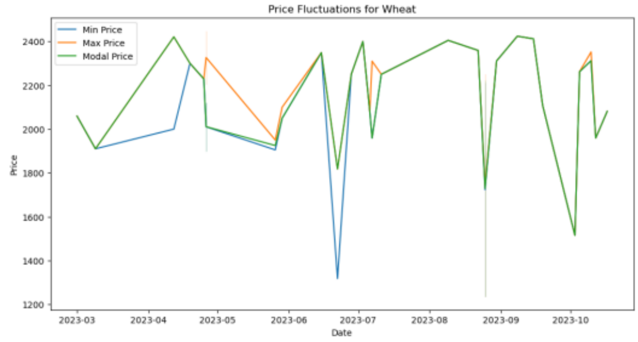
vii. The percentage of the variability in the dependent variable that can be accounted for by the single or many independent variables is referred to as the R-squared (R2) value.

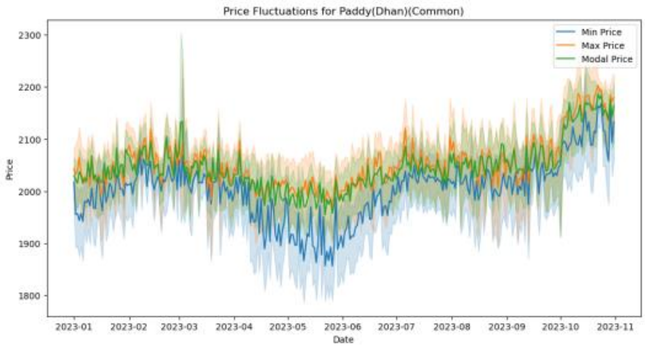
Data visualization:

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data

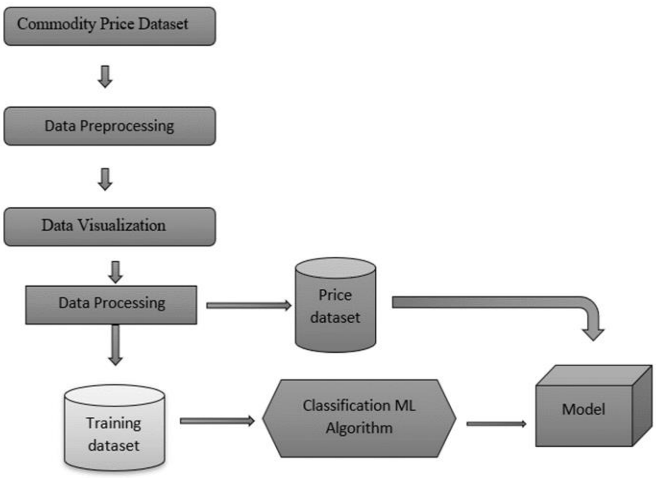
VOLUME: 08 ISSUE: 01 | JANUARY - 2024 SJIF RATING: 8.176 ISSN: 2582-3930

visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some of the books mentioned at the end.

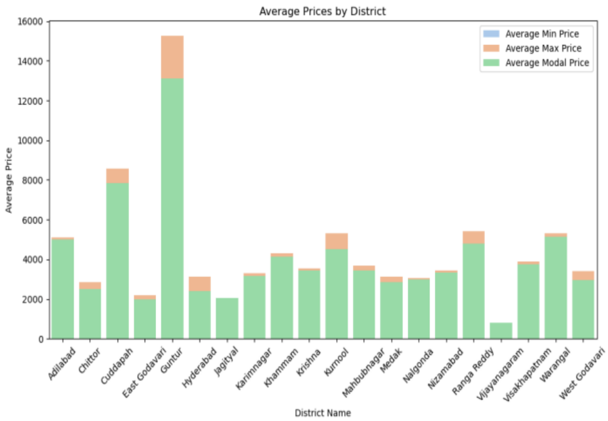


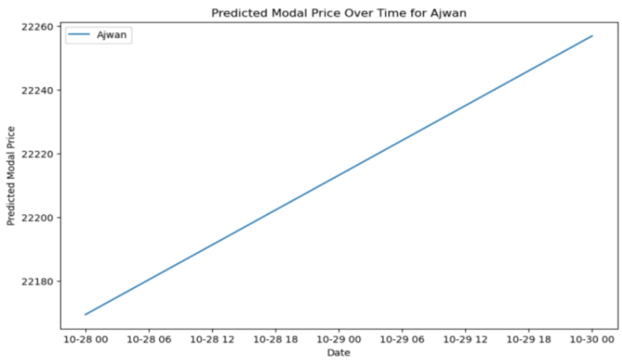


Architecture Diagram:

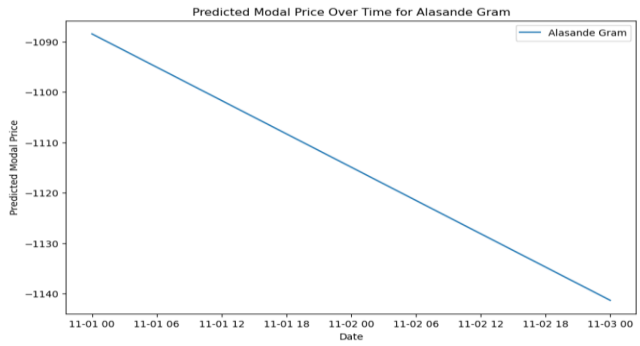


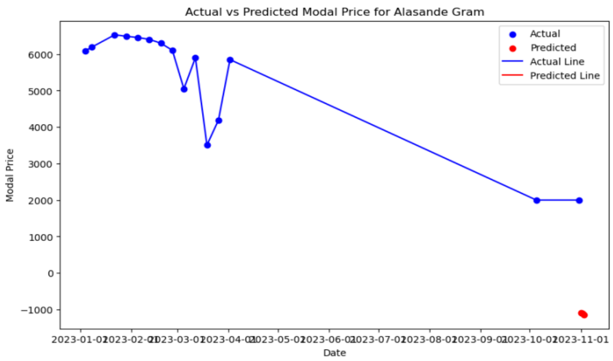
Outcomes:

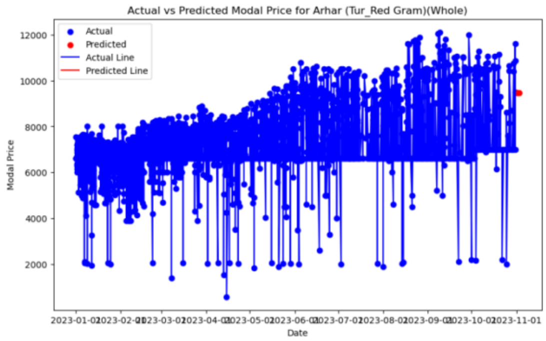




VOLUME: 08 ISSUE: 01 | JANUARY - 2024 SJIF RATING: 8.176 ISSN: 2582-3930







CONCLUSION:

- Forecasting commodity prices is a necessary dance with uncertainty, but it's hardly a kind of divination. There are some parts that remain blind, despite our efforts.
- Complicated interdependencies, hints from real-time data, and mutterings from black swans pose challenges to our models. The path forward
- Acknowledge complex data, develop models that are flexible enough to handle non-linearity, and quantify that illusive, constant fuzziness.
- Carefully inspect every item, observing its unique patterns and indications of sustainability. This isn't just about assigning a number to the future market; it's also about painting a vibrant, dynamic image of it.
- By completing these research gaps, we may ultimately establish a more resilient, sustainable dance floor for the global commodities market, manage uncertainty, and make well-informed decisions.

REFERENCES

1. Brownlee, Jason. 2017. Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future. Machine Learning Mastery.
2. Cartwright, Hugh M. 2020. Machine Learning in Chemistry: The Impact of Artificial Intelligence. Royal Society of Chemistry.
3. Chang, Fi-John, Kuolin Hsu, and Li-Chiu Chang. 2019. Flood Forecasting Using Machine Learning Methods. MDPI.
4. Financial Crisis Inquiry Commission. 2011. The Financial Crisis Inquiry Report: The Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States Including Dissenting Views. Cosimo, Inc.
5. Food and Agriculture Organization. 2019. The State of Food Security and Nutrition in the World 2019: Safeguarding Against Economic Slowdowns and Downturns. United Nations.
6. Food and Agriculture Organization of the United Nations. 2013. The State of the World's Land and Water Resources for Food and Agriculture: Managing Systems at Risk. Routledge.
7. ———. 2021. The Impact of Disasters and Crises on Agriculture and Food Security: 2021. Food & Agriculture Org.
8. International Labour Office, International Monetary Fund, OECD, The World Bank, The United Nations

Economic Commission for Europe, and Statistical Office of the European Communities, Luxembourg. 2004. Consumer Price Index Manual: Theory and Practice. International Labour Organization.

9. OECD, International Labour Office, International Monetary Fund, Statistical Office of the European Communities, United Nations, and The World Bank. 2010. Export and Import Price Index Manual: Theory and Practice. OECD Publishing.

10. Pearson, Andrew. 2017. The Predictive Retailer: Making the Retailer Smart. Createspace Independent Publishing Platform.