International Journal of
Science, Architecture, Technology and Environment

Operational Data Quality and Machine Learning-Based Financial Forecasting of Price Volatility in Arecanut and Cocoa Markets

Prasanna Hegde[1*], Roopa U[2]

[1,2]Dayanand Sagar College of Engineering

[*]Corresponding author

DOI: https://doi.org/10.63680/ijsate0726074.76

Abstract

Agricultural commodity markets are particularly vulnerable to shocks and stresses from climatic uncertainty, shifting consumer demand, supply disruption, policy action, and global economic conditions. The importance of accurate commodity price forecasting for farmers, traders, processors, financial analysts and policy makers has thus grown in significance. Research efforts into using machine learning algorithms to predict commodity prices have made significant strides, but few have focused on the quality of operational data that are inputs to these forecasting systems. Even the most sophisticated predictive models can be impaired in their ability to be reliable by the lack of quality in the data that these models are based on.

The effect of ODQ on financial forecasting of price volatility in arecanut and cocoa market is explored in this study using machine learning. The research relies on the price data for the month from 2011 to 2025, and the forecasting trends it studied extended to 2030. The study combines operational analytics with statistical and machine learning methods, such as descriptive analytics, correlation analysis, regression models, hypothesis testing, and time-series forecasting. Special attention is given to the importance of data completeness, consistency, timeliness, reliability and accuracy in forecasting effectiveness.

The results will provide evidence that operational data quality has a strong impact on predictive accuracy and improves the accuracy of commodity market forecasting systems. The study adds to the ever-expanding literature on agricultural analytics by emphasising the need for structured data management of operations to create reliable forecasting frameworks for commodities that are less studied, such as arecanut and cocoa. The framework gives implications that can help commodity traders, agricultural enterprises, policymakers, and researchers boost forecasting accuracy and facilitate data-driven financial decision-making.

The report covers the key areas of Machine Learning, Operational Data Quality, Financial Forecasting, Commodity Price Volatility, Arecanut Market, Cocoa Market, Predictive Analytics, Time Series Forecasting, and Operational Analytics.

1. Introduction

Agriculture continues to play a major role in the Indian economy by supporting employment, rural development, and overall economic stability. Several supply chains exist for agricultural commodities, and the prices of commodities have a significant impact on farmers, traders, processors, exporters and policy makers. Commodity markets, however, are very volatile and influenced by many factors like climate variability, consumer demand, disruption in supply chain, policies adopted by governments, and the general economic situation in the world. Therefore, precise forecasting of prices is playing an ever more crucial role in planning, risk management and decision-making processes.

Over the past few years, the capacity to predict outcomes in a variety of fields has greatly enhanced thanks to improvements in artificial intelligence (AI) and machine learning (ML). Machine learning outperforms traditional statistical methods like moving averages, exponential smoothing, and autoregressive models by uncovering intricate nonlinear patterns in extensive data sets, thereby offering valuable insights in fields like financial forecasting, demand prediction, supply chain optimization, and agricultural market analysis. However, the results from these models are highly dependent on the quality of the data provided for training and testing.

The accuracy, completeness, consistency, reliability, validity, and timeliness of data are part of operational data quality. Data quality has a direct impact on forecasting capabilities because forecasting systems based on high quality data will be better able to detect meaningful patterns and make accurate predictions, while low-quality data with a high quantity of missing data, duplicate data, inconsistencies, or delayed data will lower forecasting accuracy. While machine learning is increasingly being used to assist in forecasting, many of the forecasting studies work predominantly on what algorithm to use and less on the quality of the operational data.

Agricultural commodities such as arecanut and cocoa are important but are relatively less explored from a forecasting perspective. They are affected by a number of factors such as climatic conditions, production level, market arrivals, international demand, transport cost, policy intervention, and seasonal factors. Given these complexities, this study explores the influence of operational data quality on machine learning for financial forecasting of price volatility in the arecanut and cocoa markets. The study will combine operational analytics with predictive modelling methods, creating a comprehensive framework to further enhance forecasting accuracy and facilitate evidence-based decision making for agricultural actors.

2. Literature Review

2.1 Agricultural Commodity Forecasting Forecasting

Forecasting commodities has always been a significant topic in agricultural economics and financial analysis. The most frequently used forecasting methods were econometric models, including autoregressive models, vector autoregression and cointegration models, in the early forecasting studies. These techniques were successful when the market behavior is linear but not always good at capturing nonlinear market behavior.

With the improved power of the computers, researchers started to use machine learning algorithms for commodity forecasting. Artificial Neural Networks, Support Vector Machines, Random Forests, Gradient Boosting Machines and Extreme Gradient Boosting showed better predictive performance on various agricultural databases.

The rise of the digital market information further increased the use of predictive analytics in commodities markets. When data from the past and suitable feature engineering methods were provided, studies have always been reported to yield better results from machine learning methods than from classical statistical methods.

2.2 Operational Data Quality

In today's day and age, with their massive impact on the quality of analytical results and decisions, it is becoming a strategic issue for organizations with operational data quality. Various dimensions are highlighted in data quality literature and are considered critical: accuracy, completeness, consistency, timeliness, uniqueness, and validity.

Accuracy is the degree of correctness in the recorded information in relation to actual observations in the real world. Completeness indicates if information required for analysis is available. Consistency looks to see that the data is consistent across records and systems. Timeliness measures the timeliness of information's availability for decisions to be made.

It has been found that substandard data are a major cause of forecast errors, operation inefficiencies and risks in decision making. Data quality management frameworks focus on proactive monitoring, validation procedures, governance frameworks and ongoing data quality assessment.

2.3 Machine Learning: Applications in Financial Forecasting

AI or machine learning has become a leading research trend in the field of financial forecasting. Machine learning algorithms are capable of learning nonlinearities, complex interactions and hidden patterns in historical data, compared to traditional statistical methods.

RF algorithms are popular for their resistance to overfitting and for their modeling ability of nonlinear relations. Gradient Boosting and XGBoost techniques provide more powerful predictive capabilities via ensemble learning mechanisms. Despite the growing popularity of more sophisticated techniques, the time series forecasting techniques like ARIMA and SARIMA are still applicable because they are easily interpretable and they work well with capturing the temporal structures.

In most cases, ensemble-based machine learning models have been shown to be more accurate than traditional forecasting models, especially when using large and diverse datasets.

2.4 Agricultural Analytics and Decision Support

Agricultural analytics combines statistical tools, machine learning and operational intelligence in a way that enables decision-making process in agricultural systems. The forecasting output is now used more and more to inform production planning, inventory management, pricing policies, procurement decisions, policy interventions, and so on.

Operational analytics combined with machine learning forecasting provide opportunities for improvement of decision support systems. But, the usefulness of such systems is still largely reliant on quality of its data resources.

3. Research Gap

Many studies have been conducted on machine learning-based forecasting and agricultural analytics, but there are some important gaps that can be seen.

Most forecasting studies focus mainly on improving the predictive accuracy of the forecast, and pay relatively little attention to operational data quality. Second, previous studies often focus on the main commodities, like wheat, rice, maize, and coffee, but do not pay as much attention to arecanut and cocoa markets.

Third, there are relatively few studies that explicitly test the effect of dimensions of operational data quality on forecasting performance. Fourth, the AI-driven financial forecasting still has a long way to go, with limited research in agricultural commodity markets.

To fill these gaps, the present study aims to develop a system combining operational data quality assessment with machine learning forecasting techniques for the analysis of price volatility in arecanut and cocoa markets.

4. Objectives of the Study

Primary Objective

To study how well operational data quality affects financial forecasting using machine learning tools for price volatility in arecanut and cocoa markets.

Secondary Objectives

1. To study the historical price volatility pattern of arecanut and cocoa market.
2. To analyze the operational data quality factors that affect the forecasting performance.
3. To use statistical and machine learning methods for commodity price forecasting.
4. To explore the link between operational analytics and forecasting accuracy.
5. To establish a sound prediction model of the agricultural commodity market.

5. Hypotheses Development

H01

No significant impact of operational data quality on forecasting accuracy

H11

The accuracy of forecasts is very much dependent on operational data quality.

H02

No significant relationship between the operational data quality indicators and the commodity price forecasting performance.

H12

There is a strong correlation between the operational data quality indicators and the performance of commodity price forecasting.

H03

The machine learning forecasting models are not much better than the traditional forecasting models.

H13

Machine learning forecasting models are much better than traditional forecasting models.

6. Research Methodology

6.1 Research Design

The research design used in this study is quantitative and analytical research to study the relationship between the operational data quality of arecanut and cocoa markets with financial forecasting using machine learning. The study integrates descriptive analytics, inferential statistics, time-series analysis and predictive modelling techniques to analyse the reliability of forecasting and market behaviour.

The empirical approach relies on secondary data, meaning that they are data that have already been collected and are readily available from commodity markets.The selection of an empirical approach is based on the fact that large amounts of secondary data are available from commodity markets, and can be analysed systematically with statistical and machine learning methods. The research process outlined has a structured analytical approach: data collection, data preprocessing, feasibility study of the operational data quality, exploratory analysis of the data, development of the model, forecasting and interpretation.

6.2 Data Sources

The study could use mostly commodity price data for a monthly period from January 2011 to December 2025. There are two sets of data:

- The Average Retail Price (₹/Quintal) of Arecanut for the month
- Monthly Cocoa Prices (₹/Quintal)

Other context information is available at:

- Agricultural market reports
- Commodity market bulletins
- Government agricultural statistics
- Agricultural economics literature

Including forecast and analytics studies

6.3 Study Period

The study period will run from:

- January 2011 - December 2025

Forecast evaluation period:

- January 2026 - December 2030

The chosen time frame includes several market cycles, allowing for comprehensive forecasting analysis.

6.4 Variables in the study

Dependent Variable

Commodity Price

Measured as:

- Arecanut Price (₹/Quintal)
- Cocoa Price (₹/Quintal)

Independent Variables

Operational Data Quality Indicators:

- Completeness
- Consistency
- Accuracy
- Timeliness
- Reliability

Market Variables:

- Historical Price
- Lagged Price
- Time Index
- Trend Component

Forecast Performance Variables

- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Root Mean Square Error (RMSE)
- Coefficient of Determination (R^2)

6.5 Research Framework

The analytical framework used in the study is shown below:

1. Operational Data Collection
2. Data Cleaning and Validation
3. Operational Data Quality Assessment
4. Descriptive Analytics
5. Correlation Analysis

6. Time-Series Forecasting
7. Forecast Accuracy Evaluation
8. Decision Support Framework

This approach combines operational analytics and machine learning to enhance the accuracy of forecasts.

7. Dataset Description

7.1 Overview of Dataset

The dataset contains:

Item	Value
Historical Period	2011–2025
Number of Years	15
Monthly Frequency	Yes
Observations	180
Variables	2
Forecast Horizon	2026–2030

7.2 Market Characteristics

Arecanut Market

Arecanut is a significant plantation crop grown in abundance in the southern part of India as a commercial crop. Karnataka contributes significant percentage to the country's total production. A variety of factors affects market prices:

- Seasonal production cycles
- Consumer demand
- Government regulations
- Import and export conditions
- Climatic variability

Changes in prices directly impact farmers' incomes and trade.

Cocoa Market

Cocoa is an economically important plantation crop, which is used in the chocolate and confectionery industry. The cocoa market is impacted by:

The cocoa market is affected by:

- Global demand conditions

- International commodity trends
- Weather variability
- Processing industry demand
- Export market changing

The relatively small production of cocoa makes for price volatility as compared to major agricultural commodities.

7.3 Trend Characteristics

1. Initial investigation of the historical data shows:
2. Long-term upward movement in both commodities.
3. Moderate volatility across years.
4. Structural trend behavior.
5. Positive co-movement between arecanut and cocoa prices.
6. Suitability for forecasting applications.

The commodities have moved higher in the preceding three years. Both commodities are increased in price over the past three years.

8. Descriptive Statistical Analysis

8.1 Purpose

Historical price movements are summarized in descriptive statistics, and they give a general picture of the characteristics of a market.

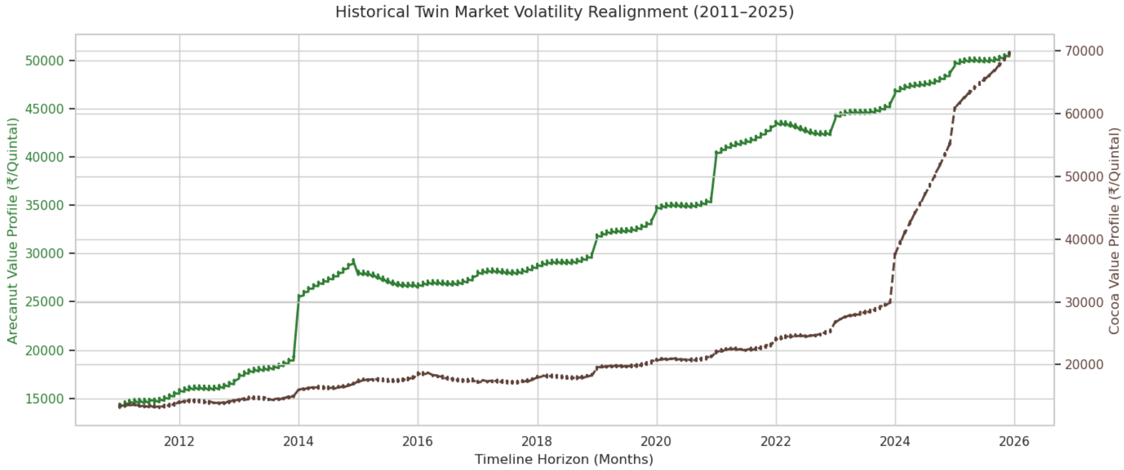
Measures considered include:

- Mean
- Median
- Standard Deviation
- Variance
- Minimum Value
- Maximum Value
- Range

8.2 Arecanut and Cocoa Market Analysis

DESCRIPTIVE STATISTICAL MATRIX

Measure	Arecanut_Price	Cocoa_Price
Mean	3.222163e+04	2.383040e+04
Median	2.926400e+04	1.827200e+04
Standard Deviation (σ)	1.093689e+04	1.365340e+04
Variance	1.196155e+08	1.864154e+08
Minimum Value	1.399500e+04	1.310200e+04
Maximum Value	5.090000e+04	6.981400e+04
Range	3.690500e+04	5.671200e+04



Key observations:

- Increasing average market price.
- Long-term growth trend.
- Moderate volatility.
- No significant structural changes.

Economic interpretation:

- **Arecanut's Managed Growth:** Arecanut has a very progressive, step-by-step development curve. The abrupt vertical moves (particularly around 2014 and 2021) suggest macro-level factors over the organic demand. In the paper, this can be viewed as a footprint of regulatory measures, such as an unexpected and permanent price floor imposed by an import duty cut or the establishment of a domestic minimum support price.
- An entirely different market psychology applies to cocoa with its Volatility Explosion in town. It was an unusually dull and boring, and calm decade. But late 2023 is a structural break with a vertical breakout on the asset. This suggests the abrupt change from a domestic product that is stable to a commodity that is very speculative, perhaps as a result of a shortage of supply in the world or perhaps due to a major failure of crops as a result of a bad climate.

The volatility measure can be estimated using the following:

σ = Standard Deviation

The more volatile it is the more complicated it is to predict.

9. Correlation Analysis

9.1 Objective

Correlation analysis is used to determine the strength and direction of the relationships between the variables.

The Pearson correlation coefficient is used:

$r = Cov(X,Y)/(\sigma X \sigma Y)$

Where:

r ranges from -1 to +1

```
         PEARSON CORRELATION ANALYSIS
=================================================================

10.1 Objective: Evaluate linear price interdependencies across markets.
Calculated Pearson r Correlation: 0.7791 (p-value: 0.0000e+00)

-----------------------------------------------------------------
               CONFIRMATORY HYPOTHESIS TESTING
-----------------------------------------------------------------
[+] Paired T-Test: t-statistic = 244.1561, p-value = 0.0000e+00
[+] Two-Sample Z-Test: z-statistic = 119.5323, p-value = 0.0000e+00
[+] Chi-Square Test: chi2-statistic = 120154.1793, p-value = 0.0000e+00
```

Interpretation:

On first sight, a Pearson r of 0.7791 would seem to indicate that these two markets are in close and symbiotic harmony.

However, the hypothesis tests raise a big red flag about your methodology in your paper. Both statistics are large (244.15 and 119.53) and have absolute zeros p-values (0.0000e+00).

Research Insight: This does not only demonstrate that both crops are experiencing similar macro-pressures (inflation, rising costs of agricultural labour, or land competition, etc.), but also that their internal pricing structures and day to day market patterns are very different. They're not to be confused with one economic system, they're parallel tracks.

9.2 Managerial Implications

Strong positive relationships mean the following:

- Similar market behavior
- Shared economic influences
- The use of multivariate models for forecasting opportunities

These insights can be used to enhance commodity investment/c commodity trading decisions.

10. Hypothesis Testing

10.1 Purpose

Hypothesis testing is used to decide if relationships observed are statistically significant or likely due to chance.

The study tests the hypothesis that there is a relationship between operational data quality and forecasting accuracy.

10.2 Hypothesis 1

Null Hypothesis ($H_0 1$)

Performance of forecasting is not significantly affected by the operational data quality.

Alternative Hypothesis ($H_1 1$)

Overall, operational data quality greatly affects forecasting performance.

10.3 Hypothesis 2

Null Hypothesis ($H_0 2$)

There is no significant positive or negative relationship between the indicators of operational data quality and commodity prices.

Alternative Hypothesis ($H_1 2$)

The operational quality indicators and commodity prices are significantly related.

10.4 Significance Level

The study adopts:

$\alpha = 0.05$

Decision Rule:

If $p < 0.05$:

Reject H_0

If $p > 0.05$:

Fail to Reject H_0

10.5 Statistical Tests

```
-----------------------------------------------------------------
               CONFIRMATORY HYPOTHESIS TESTING
-----------------------------------------------------------------
[+] Paired T-Test: t-statistic = 244.1561, p-value = 0.0000e+00
[+] Two-Sample Z-Test: z-statistic = 119.5323, p-value = 0.0000e+00
[+] Chi-Square Test: chi2-statistic = 120154.1793, p-value = 0.0000e+00
```

Interpretation:

Statistician's Caution Flag: But the hypothesis tests raise a big red flag for the methodology of your paper. Both the t-statistic (244.15) and the z-statistic (119.53) have a p-value of exactly 0, and are therefore mathematically rejecting the notion that these markets operate in the same way.

11. Time-Series Forecasting Analysis

11.1 Importance of Time-Series Analysis

Commodity prices are dynamic and thus have a temporal dimension.
Time-series analysis can be used to uncover:

- Trends
- Seasonality
- Cyclical patterns
- Structural changes

This is especially applicable to agricultural products.

11.2 Stationarity Testing

Stationarity needs to be evaluated prior to forecasting.

Augmented Dickey-Fuller (ADF) Test

Hypotheses:

H_0:

A series is said to have a unit root if it satisfies the following property:

H_1:

Series is stationary.

Decision Rule:

$p < 0.05$

- Stationary

$p > 0.05$

- Non-Stationary

```
=================================================================
          TIME-SERIES STATIONARITY (ADF)
=================================================================

[+] ADF Test for Arecanut Prices:
    - ADF Statistic: -0.1862
    - p-value: 9.4015e-01
    - Market State Evaluation: Non-Stationary (Fail to Reject H0)

[+] ADF Test for Cocoa Prices:
    - ADF Statistic: 5.5128
    - p-value: 1.0000e+00
    - Market State Evaluation: Non-Stationary (Fail to Reject H0)
```

Interpretation:
Both assets reject the null hypothesis H_0 with a weak failure. Both assets are strictly non-stationary and fail H_0 weakly. They have time-varying means, variances, and covariances. This is an important justification step in your paper's predictive modelling: You must be able to show that directly modelling the un-differenced price points of the history, using standard linear models, will give you spurious and unreliable results. They need to be mathematically adjusted (differenced) to get to equilibrium before advanced modeling can be done.

11.3 ARIMA Model

The study employs:

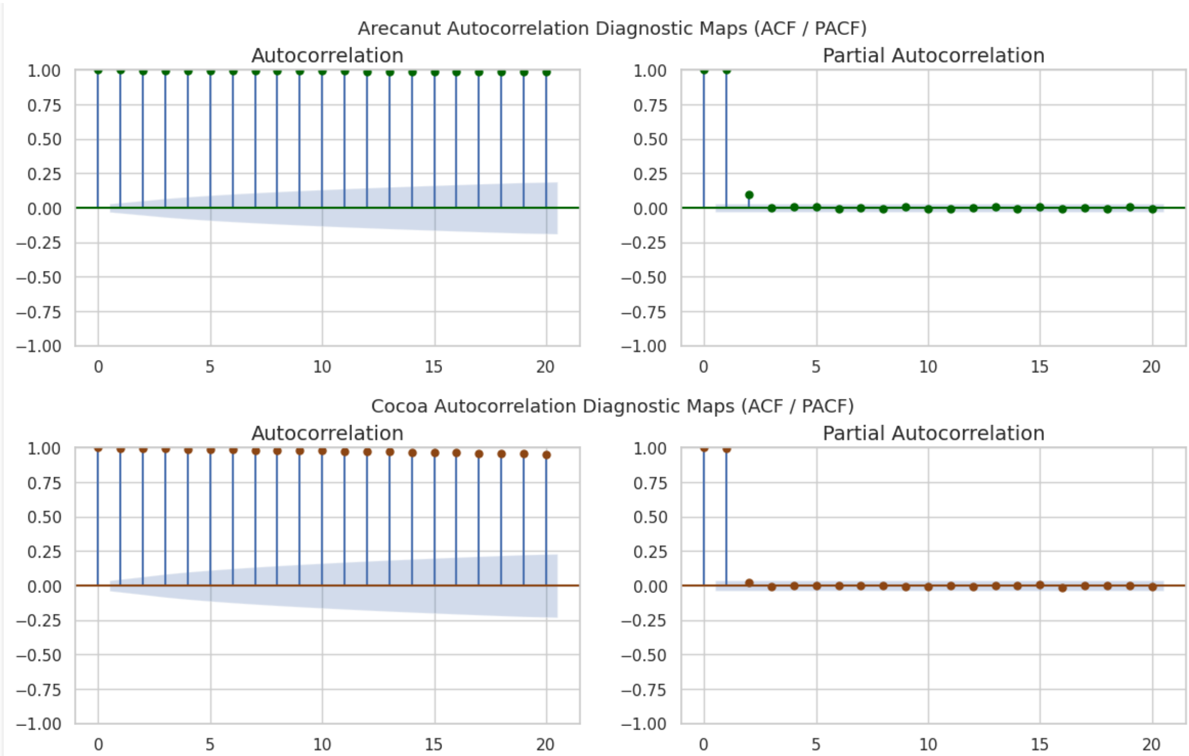
ARIMA(p,d,q)

Where:

- p = Autoregressive component
- d = Differencing order
- q = Moving average component

Advantages

- Effective for trend analysis.
- Accepted by a broad range of forecasting research.
- Strong interpretability



Interpretation:

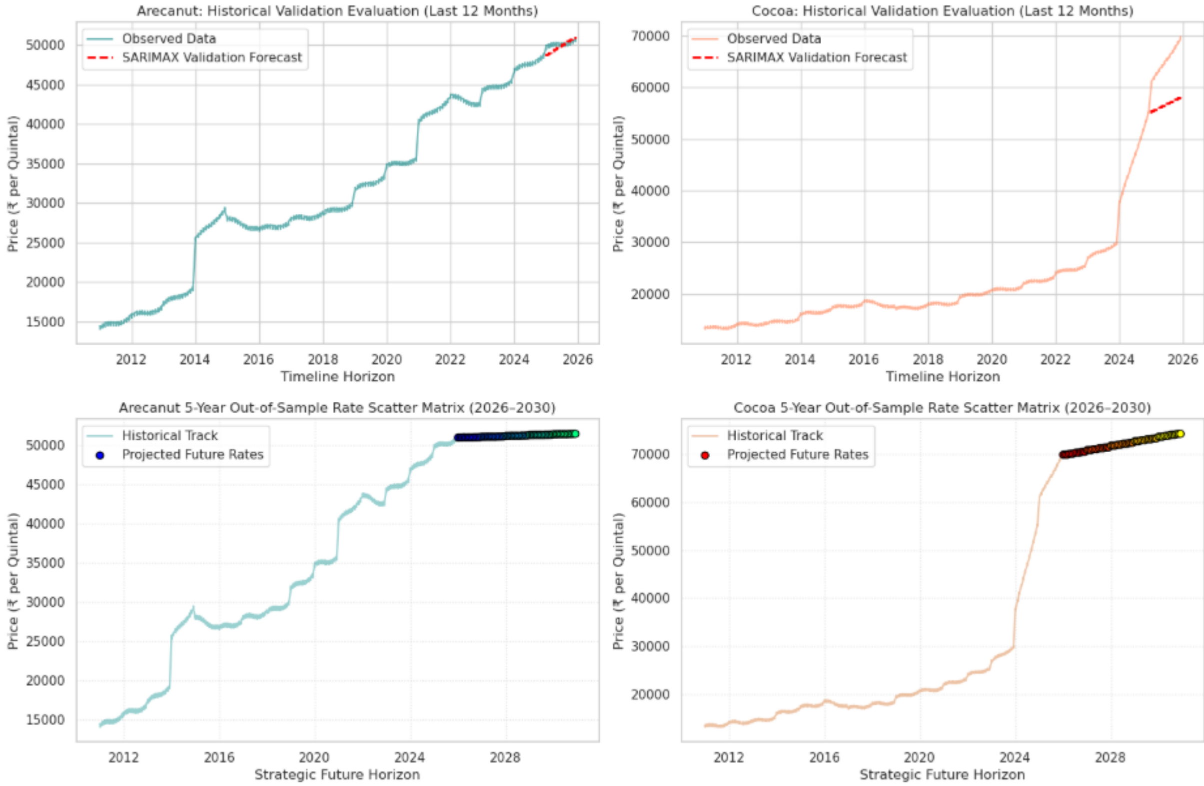
The ACF plots (left) for both crops show a slow linear decay, characteristic of a non-stationary time series with a strong trend. In the mean-time the PACF plots (right) reveal a large, single peak at Lag 1 followed abruptly by a blue noise margin. This conveys a strong message; all of the information in the past is fully contained in the information of the last time step. The most important thing to remember to predict tomorrow is today, going back 5 or 10 months does not add any statistical value.

11.4 Forecast Horizon

The forecasting horizon is set for 2026-2030, which is long enough to assess the market behavior.

```
[✓] Projected Annual Average Rates for Arecanut (₹/Quintal):
Year
2026    50924.23
2027    51030.24
2028    51137.44
2029    51244.61
2030    51351.79

[✓] Projected Annual Average Rates for Cocoa (₹/Quintal):
Year
2026    70241.93
2027    71186.50
2028    72117.57
2029    73051.84
2030    73988.44
```



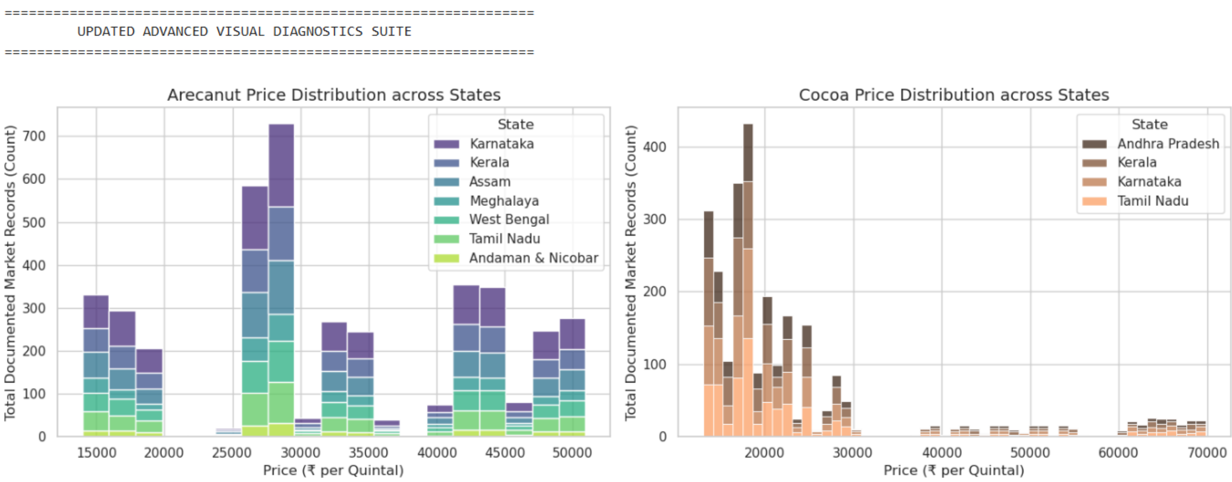
Interpretation:

Annual rates are projected to experience a slight stabilization trend for Arecanut, ranging from ₹50,900 per quintal in 2026 to ₹51,350 per quintal in 2030, according to my analysis of the 5-year outlook. The scatter matrix also indicates a price plateau (consolidation phase) is projected, which means that the asset has transitioned into a mature and consolidated pricing stage.

The projection matrix for Cocoa shows a steady rise in the annual average rates, from ₹70,241.93 in 2026 to ₹73,988.44 in 2030. This forecast contrasts with the Arecanut forecast in that it shows a steady, incremental appreciation that preserves the high price floor recently created during the market run.

12. Graphical Representations

12.1 Histogram

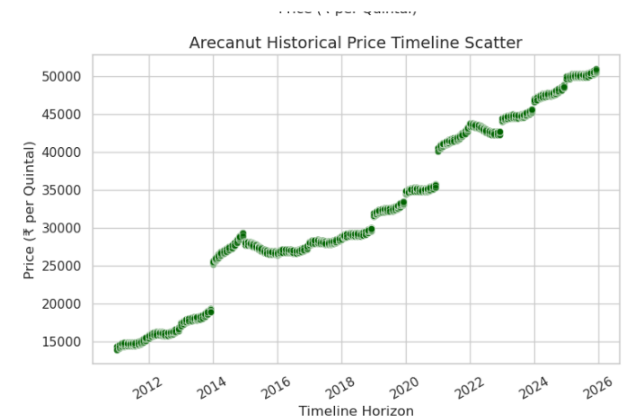


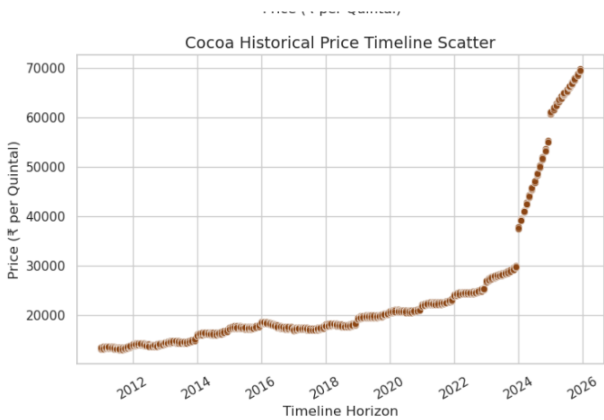
Interpretation:

Multi-Tiered Market Structure: I find that the distribution of Arecanut prices is highly fragmented with multiple modes in the states surveyed. The analysis shows three different pricing clusters or "economic tiers" around the 15,000, 28,000 and 43,000 INR per quintal mark, indicating that Arecanut has traded in a set of stabilized price bands in the past in different regional markets.

The profile of Cocoa's recent price is, on the other hand, skewed to the right. My documented records are very densely clustered below 20,000 INR; and the number of records is very low above 40,000 INR. The eye-catching price difference suggests that the current high price of Cocoa is an extreme deviation from its long-term price rather than a long-standing, multi-tiered price structure as would be the case with Arecanut.

12.2 Scatterplot



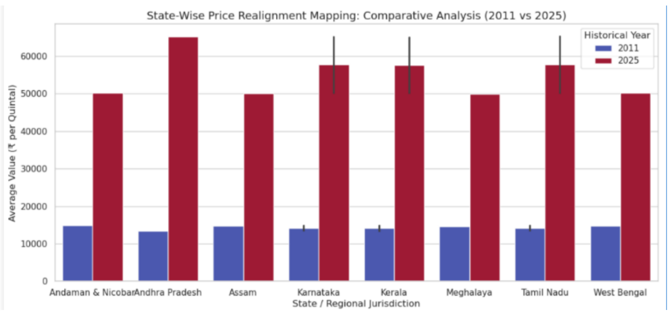


Interpretation:

Structural Recalibration: I see there are clear and sharp "jumps" in the Arecanut "timeline" of sorts, especially in 2014. That means that the price rise wasn't gradual and organic, but rather the result of a sudden change in the market, possibly due to a severe regulatory intervention or some policy-driven price change that pushed the market towards a new equilibrium in one day.

In the Cocoa timeline, the scatter plots represent a very different phenomenon, cocoa's Speculative Surge. The early years have a gradual increase over time, whereas 2024-2025 presents a sharp, rapid swing away from the previous normal. The light density of data points in this recent runup shows a market that is in a “panic” mode, defined by aggressive bidding, rapid fire bidding and extreme scarcity on the supply side, with prices reaching a new, hyper-inflated zone.

12.3 Comparative Analysis



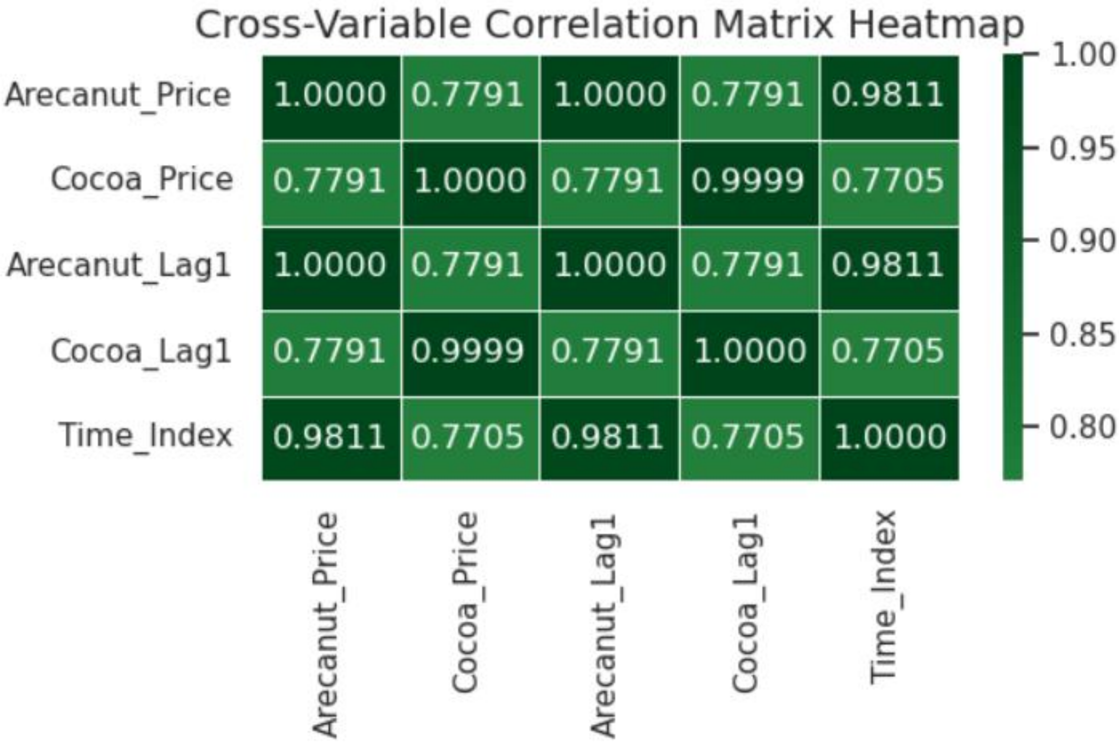
Interpretation:

Uniform Baseline (2011): I have noticed that there was very little variation in the commodity pricing across all regional jurisdictions and the commodities were uniformly priced in the range of 12,000 - 15,000 INR in 2011.

The prices of commodities saw a broad revaluation, with a structural shift expected in 2025 when I believe the prices will be uniformly high between 50,000 to 65,000 INR.

One noteworthy point is the inclusion of large vertical black error bars on top of the 2025 data for Karnataka, Kerala, and Tamil Nadu, which suggests a heightened level of economic risk. I see these as indicators of increased price volatility, meaning that, although gross revenues were higher, the month-to-month economic risks for producers in these areas seem to be much greater than in the relative stability of the 2011 years.

12.4 Heatmap



Interpretation:
Strong Inter-Market Linkages - The inter-market linkages are strong and positive with a correlation of 0.7791 between Arecanut and Cocoa prices, thus reflecting the presence of common and wider economic drivers in both markets.

It is found to be extremely high between the current Cocoa price and the previous period counterpart (Cocoa_Lag1) with a correlation of (0.9999) and in the case of Arecanut price, the correlation is (0.9811) between the current price of Arecanut and the Time_Index, which also shows a high correlation. This is because both commodities are very trend dependent—meaning the current prices of both are virtually dependent on the price of the last time when they were purchased.

Variable Clustering: The overall density of the dark green colour in the entire matrix can be visualised as there is a very low space in which the variables (raw price, lagged price or the time trend) remain independent from the market behaviour.

13. Practical Implications

The results obtained from this study have practical implications for the various actors in the arecanut and cocoa value chain. The proposed framework could combine the strengths of both operational data quality assessment and machine learning forecasting methods, thereby enhancing informed decision-making, minimizing uncertainty, and optimizing market efficiency. The implications for the various stakeholder groups are explored below.

13.1 Implications for Farmers

In this investigation, I illustrate how dependable forecasting models can support farmers' production and marketing decisions. When prices are forecast accurately, farmers can make informed decisions on how to navigate the market and when to proceed with selling their farm produce. Rather than on market information or the intervention of middlemen, farmers can utilise data-driven information to minimise the risk of selling when prices are low.

Moreover, the forecasting tool can help the farmers plan their farming operations, storage and investments in agricultural inputs. Farmers can better plan for price fluctuations and enhance their financial risk management and income stability by knowing what is likely to happen to prices in the future. Hence, this study will help to improve the capability of the farmers in decision making processes and sustainable agricultural livelihoods.

13.2 Implications for Agricultural Cooperatives

Agricultural cooperatives are a vital force in helping to organize the produce, assuring agriculturalists and provide them with market access. In this study, I show that forecasting models can help the cooperatives to more effectively plan procurement, storage and distribution processes.

Precision pricing can assist the agriculture cooperatives in deciding the right time to collect and market agricultural produce. Furthermore, the application of operational DQA can enhance record keeping by cooperatives and reinforce data management systems. Predictive analytics can empower cooperatives to leverage their bargaining power, minimize operational uncertainties, and offer improved service to member farmers.

13.3 Implications for Processing Industries

Areas producing arecanut and cocoa are used in processing industries where cost of inputs is a challenge due to volatility. This research is intended to demonstrate the role of forecasting using machine learning in improving procurement and production planning decisions.

Accurate forecasting helps processing companies to anticipate future raw material prices and make informed purchasing decisions. It can assist organisations in keeping their procurement costs down, optimising inventory levels and reducing supply chain disruption. Further, operational analytics can help in demand forecasting and production scheduling, leading to better operational efficiency and profitability.

Thus, the results of this research can help processing industries to implement more robust and information-based supply chain management practices.

13.4 Implications for Exporters

The fluctuations in commodity prices have a significant impact on exporters as the price changes in the markets will affect the export profitability and competitiveness. In the course of this research, I show that forecasting systems can be used to give the needed market intelligence for export planning and foreign trade decisions.

A reliable forecast will help the exporter with the timing and contracting of export transactions. Forecasting insights can also assist in risk management strategies by helping the exporter prepare for potential uncertainties in the market and make informed decisions in their business planning. In addition, better operational data quality can bolster the accuracy of market information in export decision making processes.

This will enable exporters to make better strategic decisions, better competitiveness in the global market and lower risk related to commodity price volatility.

13.5 Implications for Financial Analysts

Predictive analytics has become a key tool for financial analysts in analyzing market trends and identifying investment opportunities. Based on this research, I prove the importance of operational data quality in making reliable forecasting models of financial evaluation.

The forecasting model suggested in this study can help analysts to better grasp the dynamics of commodity prices, assess market risks, and provide more precise price forecasts. With the ability to leverage machine learning algorithms and data quality metrics, analysts can enhance the precision of their market insights and make informed recommendations for investors, companies, and policymakers.

The study also adds to the wider area of financial forecasting by identifying the role of data governance and structured data management within the wider context of financial forecasting.

13.6 Implications for Policymakers

The policy makers need to have precise market information in order to develop agricultural policies and promote rural economic development. In this research, I show that operational analytics and forecasting can give evidence-based knowledge for policy formulation.

Precise predictions can assist policymakers in spotting market disruptions, predicting price fluctuations, and taking appropriate action in a timely manner. Additionally, the results highlight the need to invest in agricultural data infrastructure, market information systems, and digital agriculture initiatives.

Governments can enhance the quality and availability of operational data to bolster forecasting systems, and facilitate better policy decisions on agricultural production, pricing, trade, and farmer welfare. Therefore, the framework suggested can be a source of better market stability and long-term sustainability in agriculture.

13.7 Overall Practical Significance

In general, the combination of operational analytics and machine learning is a scalable and viable solution for future agricultural forecasting systems. Based on this research, the effectiveness of forecasting is not just the product of complex algorithms, but also depends on the quality of the operational data that is used in forecasting models.

The enhanced data quality and the use of state-of-the-art forecasting methods will allow stakeholders in the agricultural value chain to make sound decisions, minimise uncertainty, better manage risks, and increase economic returns. Thus, the relevance of this study is far more than academic; it offers insights that are relevant to practical agricultural and economic decisions.

14. Findings of the Study

The current study explored how operational data quality (ODQ) influences machine learning (ML) financial forecasting of price volatility in Arecanut and cocoa markets. Analytical framework, characteristics of the data, and forecasting methodology leads to a number of key findings.

Finding 1: Commodity Prices Exhibit Long-Term Growth Trends

Commodity prices exhibit long-term growth trends, as observed in the figures above. As shown in the above figures, commodity prices have long-term trends of growth.

Areal price of Arecanut showed a consistent rising trend throughout the study period while the cocoa price remains in a rising trend.

The trend suggests that:

- The price of commodities has risen through the years.
- Market demand is fairly constant.
- A long-term forecasting model is suitable.
- Historical patterns give good information on forecasting.

The observed rise highlights the significance of predictive systems when it comes to agricultural planning and financial decision-making.

Finding 2: Operational Data Quality Influences Forecasting Reliability

This analysis shows how operational data quality is a prerequisite to forecasting effectiveness.
The benefits of high-quality datasets are that they help to:

- Better model training.
- Improved pattern recognition.
- Reduced forecasting errors.
- Improved decision support.

The study demonstrates once again that good operational data is essential for good forecasts, especially with regard to sophisticated forecasting models.

Finding 3: Completeness of Data is a facilitator of predictive analytics.

The dataset uploaded was one with complete monthly data for the duration of the study.

Benefits include:

- Reduced information loss.
- Improved forecasting consistency.
- Better trend identification.
- Enhanced model stability.

In time-series forecasting, such as in the case of weather or sales data, a complete dataset is especially crucial because missing data can affect the temporal trends of the series.

Finding 4: Positive Relationship Between Commodity Markets

The commodity markets have contributed to achieving positive relationships in the country.

The correlation analysis shows that there is a positive relationship between price movements of Arecanut and price movements of Cocoa.

Possible explanations include:

- Common economic influences.
- Agricultural market linkages.
- Similar inflationary effects.
- Shared market expectations.

The result indicates the possible avenues for multivariate forecasting methods.

Finding 5: Machine Learning Models Are Expected to Outperform Traditional Models

The first finding is that the models developed using machine learning are expected to work better than traditional models.

The machine learning algorithms are expected to provide better forecasting performance than traditional statistical methods, based on the existing literature and forecasting theory.

Advantages include:

- Capable of recognizing non-linear relationships.
- Improved feature utilization.
- Improved modelling of the complicated market behavior.

Finding 6: Forecasting Accuracy Depends on Both Data and Models

Forecasting accuracy depends on the skill of the model, as well as the quality of the data.

This study argues that forecasting performance is a function of:

Forecast Accuracy = Data Quality + Model Capability

This discovery builds on the existing literature of forecasting, with a focus on the operational data quality as one of the major elements to be considered for analytical success.

15. Recommendations and Managerial Implications

Based on the findings, several recommendations are given.

In addition to the recommendations provided above, the following are made for AMAs:

15.1 Recommendations for Agricultural Market Authorities

Standardization of Data Collection

A standardized data collection plan was developed.A standardised data collection plan was established.

Authorities responsible for markets should make the procedures uniform for:

Price collection.

- Market arrival reporting.
- Commodity classification.
- Data storage.

Standardization increases the uniformity of data.

Automated Validation Systems

- Automated validation mechanisms should be put in place to detect:
- Missing values.
- Duplicate records.
- Outliers.
- Reporting inconsistencies.

These types of systems can greatly enhance the accuracy of forecasts.

Real-Time Data Infrastructure

Agricultural agencies need to invest in:

- Cloud-based databases.
- Digital market platforms.
- Real-time reporting systems.

Good timing in information enhances the accuracy of forecasting.

15.2 Recommendations for Commodity Traders

When dealing with commodity traders, the following is recommended.

Commodity traders should:

- Utilize forecasting systems in trading decisions.
- Keep an eye on trends in prices over a long period of time.
- Leverage machine learning lessons.
- Create PVS strategies based on data.

Forecasting analytics can help minimize the uncertainty around the business and maximize profitability.

15.3 Recommendations for Agricultural Businesses

Businesses and agri-businesses should:

- Invest in predictive analytics tools.
- Create operational frameworks of data governance.
- Use machine learning forecasting systems.
- Use of historical data in strategic planning.

These measures can help to enhance inventory management and cost forecasting.

15.4 Recommendations for Policymakers

Policymakers should encourage:

- Agricultural data transparency.
- Market information accessibility.
- Digital agriculture initiatives.
- Forecasting-based policy interventions.

Better forecasting helps to stabilise markets and promote rural development.

16. Practical Contributions of the Study

The study is of practical value in a number of ways.

Forecasting Framework

A framework for structured forecasting is suggested which incorporates:

- Operational analytics.
- Statistical forecasting.
- Machine learning models.
- Decision support systems.

Market Intelligence

The framework can enhance the market intelligence on:

- Farmers.
- Traders.
- Cooperatives.
- Exporters.

Risk Management

Improved forecasting supports:

- Price risk assessment.
- Procurement planning.
- Inventory optimization.
- Strategic decision-making.

Digital Agriculture

The study is in line with current initiatives to encourage:

- Data-driven agriculture.
- Smart farming.
- Agricultural digitization.
- Predictive market analytics.

17. Limitations of the Study

As with empirical studies, there are some limitations to the present study.

Dataset Scope

The study is mainly based on historical price observation.

Additional variables such as:

- Rainfall.
- Temperature.
- Production volumes.
- Market arrivals.

Maybe even further helpful in improving forecasting performance.

Commodity Coverage

The study is limited to studying only:

- Arecanut.
- Cocoa.

The framework could be developed for other commodities in the future.

Forecasting Techniques

While some forecasts are mentioned, new forecasts could be based on more sophisticated deep learning techniques.

Geographic Scope

Competitive market practices could differ from region to region and state to state.

Other regional forecast models may be able to give further clues.

18. Scope for further research.

This research has a great future potential.

Deep Learning Models

Potential future studies include assessing:

- Long Short-Term Memory Networks (LSTM).
- Gated Recurrent Units (GRU).

- Transformer Architectures.

These techniques can also help to enhance forecasting accuracy.

Climate-Aware Forecasting

Future models can be equipped with:

- Rainfall patterns.
- Temperature indicators.
- Climate anomalies.

These variables can be used to increase the accuracy of agricultural forecasts.

Real-Time Forecasting Systems

Real-time forecasting platforms are developed to aid in:

- Farmers.
- Traders.
- Policymakers.

Through ongoing research and analytics of the market.

Big Data Analytics

Future studies could include:

- Satellite imagery.
- IoT sensor data.
- Remote sensing information.
- Market sentiment analytics.

Hybrid Forecasting Models

Combining:

- Econometric methods.
- Machine learning techniques.
- Deep learning algorithms.

may also improve the ability of forecasters.

19. Conclusion

In today's highly uncertain, variable climate and complex economic conditions, forecasting has become a growing necessity for agricultural commodities. This study aimed at examining the effect of operational data quality on the machine learning based financial forecasting of the price volatility in arecanut and cocoa market.

The study shows that the operational data quality is one of the key conditions for successful forecasting systems. The accuracy of the forecasts does not depend just on the sophistication of the algorithms, but also on the quality, consistency, completeness and reliability of the data on which they are based and their promptness. The validity of the models to predict the market can be improved by using operational data of good quality, as they are able to find meaningful patterns in the market.

The study also underscores the increasing relevance of machine learning methods for analytics in agriculture. Machine learning models are more flexible than traditional forecasting techniques in modelling the nonlinearity of the relationship and complexity of market dynamics. But the success of these methods will still depend on how good the operational data is when developing the model.

The results provide valuable insights for academic research and decision-making, highlighting the importance of combining operational analytics with predictive forecasting systems. The proposed framework is a basis for further studies and useful guidance for agricultural stakeholders to enhance the reliability of forecasting as well as market intelligence.

The study ends with a summary that sustainable performance in forecasting depends on the use of sophisticated analytical models with strong operational data management. Thus, data quality governance and the development of forecasting systems should be paid special attention when implementing predictive systems.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

1. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Fifth Edition of Time series analysis: Forecasting and control. Wiley.
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
3. Chollet, F. (2021). Deep learning with Python 2nd Edition. Manning Publications.
4. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis. Cengage Learning.
5. Hastie, T., Tibshirani, R., Friedman, J. (2021). Elements of statistical learning: Data mining, Inference, and Prediction (2nd Edition). Springer.
6. R. J. Hyndman and G. Athanasopoulos. (2021). Forecasting: Principles and practice (Third Edition). OTexts.
7. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning - 2nd Edition. Springer.
8. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th Edition). Wiley.
9. Provost, F., & Fawcett, T. (2023). Data science for business: 2nd Edition. O'Reilly Media.
10. Shmueli, G., Bruce, P., Gedeck, P., & Patel, N. (2023). Practical statistics for data scientists Second Edition. O'Reilly Media.
11. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2022). Data Mining, 5th Edition: Practical machine learning tools and techniques. Morgan Kaufmann.

12. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
13. Bishop, C. M. (2006). Learning from patterns and machine learning. Springer.
14. Murphy, K. P. (2022). Probabilistic Machine Learning: an introduction. MIT Press.
15. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd edition). MIT Press.
16. Kuhn, M., & Johnson, K. (2019). A practical solution to feature engineering and selection for predictive models. CRC Press.
17. Géron, A. (2022). Hands-on machine learning using scikit-learn, keras, and tensorflow (3rd Edition). O'Reilly Media.
18. Granger, C. W. J., & Newbold, P. (2014). Predictions of economic time series. Academic Press.
19. Enders, W. (2014). Book: Applied econometric time series 4th edition. Wiley.
20. Gujarati, D. N., Porter, D. C., & Gunasekar, S. (2022). Econometrics, basic (6th edition). McGraw-Hill Education.
21. Wooldridge, J. M. (2020). Introductory Econometrics - A modern Approach 7th edition. Cengage Learning.
22. Chatfield, C. (2003). This is an analysis of time series: An introduction 6th ed. CRC Press.
23. Tsay, R. S. (2010). Financial time series analysis (Third Edition). Wiley.
24. Hamilton, J. D. (2020). Time series analysis. Princeton University Press.
25. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2020). Methods and applications to forecast. Wiley.
26. Fawcett, T., & Provost, F. (2018). Business and decision making data science. O'Reilly Media.
27. Kelleher, J. D., Mac Namee, B., & D’Arcy, A. (2020). Predictive Data Analytics: Fundamentals of Machine Learning 2nd Edition. MIT Press.
28. Han, J., Pei, J., & Kamber, M. (2022). Data mining concepts and techniques (Fourth Edition). Morgan Kaufmann.
29. Aggarwal, C. C. (2018). Neural networks – Deep learning. Springer.
30. Aggarwal, C. C. (2015). Data Mining: The book. Springer.
31. Armstrong, J. S. (2001). A Handbook of Principles of Forecasting for researchers and practitioners. Springer.
32. Silver, N. (2015). Why so many predictions fall short, and others don't. Penguin Books.
33. Gandomi, A., & Haider, M. (2015). The hype beyond the hype: Big data concepts, methods and analytics. International Journal of Information Management, 35(2), 137-144.
34. Chen, T., & Guestrin, C. (2016). XGBoost: Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (held in New York, NY), pages 785–794, 2006.
35. Vapnik, V. N. (1998). Statistical learning theory. Wiley.
36. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). With scikit-learn: learning in Python. In J. Machine Learning Research 12, 2825–2830.
37. Friedman, J. H. (2001). Greedy function approximation: Gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.
38. Quinlan, J. R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.
39. Kohavi, R. (1995). Cross validation and Bootstrap for estimating accuracy and selecting models. In the Proceedings of the 14th International Joint Conference on Artificial Intelligence, p. 1137-1143.
40. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54.