AI-ML Driven Predictive Analytics for Forecasting Prices of Agricultural Commodities

[1]Dr. P. Ravinder Rao, [2]Neerati Saiprakash, [3]Bijja Saikrishna, [4]Gone Akshay

[1]Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.
[2,3,4]UG Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.

Abstract. Agricultural commodity price forecasting plays a pivotal role in empowering farmers, traders, policymakers, and stakeholders by enabling data-driven decisions that improve planning, reduce risk, and enhance market efficiency. This project introduces an AI and Machine Learning (ML)-powered predictive analytics system designed to forecast agricultural commodity prices by integrating diverse data sources such as historical price records, weather conditions, and market trends. Developed using Python, the system leverages Scikit-learn for robust model development and Streamlit to provide an intuitive and interactive user interface. The core objective is to build accurate and reliable forecasting models that can dynamically predict future prices based on user-provided parameters. To achieve this, multiple regression algorithms—including Linear Regression, Random Forest Regressor, and Support Vector Machines (SVM)—are explored, allowing comparative performance analysis to identify the most effective model. These models are trained on comprehensive datasets that reflect various influencing factors on commodity prices, thus enhancing the model's predictive power. After training, the selected model is serialized using Pickle, facilitating efficient storage and rapid deployment without the need to retrain. This serialized model can be seamlessly integrated into the Streamlit-based web application, where users can input relevant variables such as past prices, rainfall, temperature, or market index values to obtain real-time price predictions. The modular design ensures scalability, while the user-friendly interface democratizes access to advanced ML tools, making it practical for non-technical users in the agricultural domain. Moreover, the system's ability to simulate different scenarios by adjusting input parameters offers valuable insights for planning crop cycles, managing inventory, and optimizing pricing strategies. Ultimately, this AI-ML solution contributes to sustainable agricultural development by reducing the uncertainty associated with commodity price fluctuations. It also supports strategic decision-making and financial planning, which is critical in a sector that is often affected by unpredictable environmental and economic factors. By combining data science techniques with real-world agricultural challenges, the project not only demonstrates the power of machine learning in applied economics but also provides a functional tool that has the potential to significantly improve the livelihood of those dependent on agriculture.

Keywords: Machine Learning, Predictive Analytics, Agricultural Commodities, Price Forecasting, Python, Streamlit, Scikit-learn, Pickle

INTRODUCTION

Agriculture has always been a cornerstone of human civilization, providing essential resources such as food, raw materials, and employment for a significant portion of the global population. In many countries, particularly those with developing economies, agriculture continues to play a vital role in sustaining livelihoods, ensuring food security, and contributing to national GDP. However, the agricultural sector is inherently vulnerable to a variety of risks, including unpredictable weather patterns, pest outbreaks, global market fluctuations, and political or trade-related uncertainties. Among these challenges, the volatility of agricultural commodity prices remains one of the most critical concerns for stakeholders throughout the supply chain—from smallholder farmers to policymakers and multinational agribusinesses.

Accurate forecasting of agricultural commodity prices is essential for informed decision-making. For farmers, it helps determine what crops to plant, when to harvest, and whether to sell immediately or store produce. For traders and distributors, price predictions influence purchasing strategies and inventory management. Policymakers and government agencies rely on price forecasts to design subsidy programs, stabilize markets, and ensure food affordability. Thus, having a reliable and dynamic forecasting system can substantially mitigate risks and contribute to economic resilience in the agricultural sector.

Traditionally, agricultural price forecasting has relied on econometric models and statistical tools such as time series analysis, ARIMA models, and moving averages. While these methods have proven useful to some extent, they often fall short when dealing with non-linear relationships and complex interactions among multiple variables. The rapidly evolving data landscape and the availability of diverse data sources—including satellite imagery, real-time weather updates, global commodity indices, and social media sentiment—necessitate more advanced and adaptable forecasting techniques. This is where Artificial Intelligence (AI) and Machine Learning (ML) technologies offer a promising alternative.

Machine Learning, a subset of AI, excels in identifying hidden patterns within large and complex datasets. ML models can learn from historical data, recognize trends, and make accurate predictions without being explicitly programmed for each scenario. Unlike traditional models that may require rigid assumptions about data distribution or linearity, ML approaches are inherently flexible and scalable. This adaptability makes them particularly suitable for agricultural commodity price forecasting, where price behavior is influenced by a wide range of interdependent factors.

This project focuses on developing a predictive analytics system that leverages AI and Machine Learning to forecast the prices of agricultural commodities. The goal is to design a solution that is both technically robust and user-friendly, enabling end-users—including farmers, traders, and policymakers—to access reliable price forecasts and make data-informed decisions. The system is developed using Python, a versatile programming language widely used in the data science community, and incorporates the Scikit-learn library for implementing various ML algorithms. For the user interface, Streamlit is utilized to create an interactive and accessible web application, allowing users to input relevant data and obtain predictions in real time.

The forecasting system is built using a structured pipeline that includes data collection, preprocessing, model training, evaluation, serialization, and deployment. Historical commodity price data is combined with auxiliary variables such as weather information, seasonality, supply-demand patterns, and market indices. These features are preprocessed to ensure consistency, remove noise, and fill missing values. The processed dataset is then used to train different regression models, including Linear Regression, Random Forest Regressor, and Support Vector Machines (SVM). These models are chosen for their complementary strengths—Linear Regression offers interpretability, Random Forest captures complex nonlinear relationships, and SVM provides robust generalization with fewer data points.

Model performance is evaluated using standard regression metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) to identify the most accurate and reliable model. Once the best-performing model is selected, it is serialized using Pickle, a Python library that allows for saving and loading trained models without retraining. This serialized model is then integrated into the Streamlit-based application, where users can interact with it dynamically by providing input variables relevant to their context.

The user interface is designed to be intuitive and accessible even to users with minimal technical knowledge. It enables users to select a specific commodity, enter parameters such as recent market prices, weather forecasts, or regional factors, and receive an instant prediction of future prices. The interactive dashboard also visualizes historical trends and model performance metrics, providing additional context to support user decision-making.

One of the key advantages of this AI-ML-based system is its adaptability. As new data becomes available—whether from updated price records, changing climate conditions, or emerging global events—the models can be retrained periodically to maintain forecasting accuracy. This ensures that the system remains relevant and continues to offer actionable insights over time. Moreover, the modular design of the application allows for the incorporation of additional features or the replacement of models with more advanced algorithms in the future, such as Long Short-Term Memory (LSTM) networks or Gradient Boosting Machines (GBM).

Beyond the immediate use case of commodity price forecasting, the broader implications of this project are significant. It demonstrates how AI and ML can be harnessed to tackle real-world challenges in agriculture—a sector that has traditionally lagged in digital transformation. By reducing the uncertainty associated with price fluctuations, this system can contribute to better financial planning, reduce post-harvest losses, and improve supply chain efficiency. It also empowers farmers with the knowledge to negotiate better prices and plan their activities more effectively, thereby enhancing their overall economic well-being.

Furthermore, the system aligns with global initiatives aimed at promoting smart agriculture and sustainable development. As climate change and population growth place increasing pressure on food systems, leveraging data-driven tools becomes essential to optimize agricultural practices, reduce waste, and ensure equitable food distribution. This project provides a blueprint for integrating modern technology into agriculture in a practical and impactful way.

In conclusion, this project represents a significant step toward modernizing agricultural decision-making through AI and Machine Learning. By creating a predictive analytics system that is accurate, adaptable, and user-friendly, it addresses a critical need in the agricultural sector. The integration of diverse data sources, the application of advanced ML algorithms, and the deployment of an interactive web interface collectively contribute to a powerful tool that supports risk reduction, enhances planning, and promotes sustainable growth. As data availability and computational capabilities continue to expand, such AI-driven solutions will play an increasingly central role in shaping the future of agriculture.

LITERATURE SURVEY

1. Bhardwaj et al. (2023)

Title: *An innovative deep learning based approach for accurate agricultural crop price prediction*

Summary:

This study introduces a novel hybrid deep learning framework that combines Graph Neural Networks (GNNs) and Convolutional Neural Networks (CNNs) for crop price prediction. The authors emphasize the importance of spatial and temporal data integration, using geospatial dependencies and local market dynamics. Their model significantly outperforms traditional regression models and even some deep learning baselines by at least 20%. The architecture allows for enhanced feature extraction and efficient time-series forecasting, particularly for crops whose prices are influenced by region-specific factors.

2. Zhang et al. (2025)

Title: *Avocado price prediction using a hybrid deep learning model: TCN-MLP-Attention architecture*

Summary:

Zhang et al. propose a three-component hybrid model integrating Temporal Convolutional Networks (TCNs), Multi-Layer Perceptrons (MLPs), and attention mechanisms to forecast avocado prices. This approach captures sequential patterns, long-term dependencies, and feature relevance more effectively than single-model architectures. The inclusion of the attention layer ensures that the model identifies the most influential input variables dynamically, improving performance in volatile markets. The authors demonstrate that this method generalizes well across different forecast horizons.

3. Jain et al. (2020)

Title: *A framework for crop price forecasting in emerging economies by analyzing the quality of time-series data*

Summary:

Focusing on the quality of input data, Jain et al. offer a methodological framework to assess and improve the time-series datasets before using them in predictive models. They argue that in emerging economies, data inconsistencies and missing entries often degrade model performance. The paper proposes preprocessing techniques such as interpolation, noise filtering, and temporal regularization. Once cleaned, the dataset is applied to both ARIMA and ML models, with significant improvements in accuracy observed.

4. Chen et al. (2021)

Title: *Automated agriculture commodity price prediction system with machine learning techniques*

Summary:

Chen et al. developed a complete prediction system using ML models like Random Forest, XGBoost, and LSTM for forecasting various commodity prices. The system features an automated data pipeline and a user-friendly interface. Among the tested models, LSTM showed the best performance due to its capability to handle temporal dependencies. The paper also underscores the benefits of integrating market signals, weather data, and supply-chain indicators into the model input.

5. Jaiswal et al. (2023)

Title: *Agricultural commodity price prediction using Long Short-Term Memory (LSTM) based neural networks*

Summary:

This study applies LSTM networks to predict the price of maize in India, showing that LSTM models effectively handle seasonality and noise in price data. Their model is trained on historical price data, along with climatic variables, to enhance prediction accuracy. Compared to traditional regression models, the LSTM model offers superior Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics. The research supports the growing adoption of deep learning in agribusiness analytics.

6. R. L., M. (2021)
Title: *Forecasting spot prices of agricultural commodities in India: Application of deep-learning models*
Summary:
This paper investigates various deep learning techniques—especially LSTM and GRU—for forecasting spot prices of rice and wheat in India. The study compares these models with ARIMA and finds that deep learning models consistently outperform in terms of long-term prediction stability. A key contribution is the use of real-time market and weather inputs, enabling near-real-time forecasting applicable for farmer advisory systems and procurement planning.

7. Bhaskara et al. (2023)
Title: *Predicting prices of cash crop using machine learning*
Summary:
This paper presents a comparative analysis of ML models including Linear Regression, Random Forest, and Gradient Boosting for predicting the price of cash crops like cotton. The authors find that tree-based ensemble methods such as Random Forest deliver better performance due to their non-parametric nature and ability to manage nonlinear patterns in agricultural data. The study also stresses the importance of feature selection in boosting model performance.

8. Tran et al. (2023)
Title: *Predicting agricultural commodities prices with machine learning: A review of current research*
Summary:
This review article synthesizes recent advancements in ML applications for agricultural price forecasting. It categorizes models into statistical, machine learning, and deep learning approaches, and highlights the increasing popularity of hybrid and ensemble models. The authors discuss common challenges such as data scarcity, model interpretability, and deployment limitations. Key recommendations include developing explainable AI (XAI) systems and enhancing multi-modal data fusion.

9. Manogna, R. L. (2021)
Title: *Forecasting prices of agricultural commodities using machine learning for global food security*
Summary:
Manogna focuses on the strategic importance of price forecasting in ensuring food security. Using regression models like SVR, Ridge Regression, and Random Forest, the paper evaluates short-term price movements for key staples such as pulses and cereals. Results indicate that SVR performs best when trained with weather, import-export, and historical pricing data. The research supports policy-level use of ML for subsidy planning and market stabilization.

10. Kumar, V., Singh, A., & Rani, R. (2022)
Title: *Agricultural product price forecasting using hybrid machine learning models*
Summary:
This work proposes a hybrid ensemble framework that combines the strengths of ARIMA, LSTM, and SVR. Each model captures different aspects of the price time series: ARIMA models linear trends, LSTM captures sequential dynamics, and SVR addresses short-term nonlinearity. Results demonstrate that this multi-model approach achieves the lowest prediction errors across multiple crops. The authors also implement feature importance techniques to identify key driving variables.

11. Ali, S., Yousaf, M., & Shahbaz, M. (2021)
Title: *Machine learning to predict grains futures prices*
Summary:
This paper focuses on forecasting futures prices for grains using ML models such as Support Vector Regression (SVR), Decision Trees, and ensemble techniques. It reveals that futures markets exhibit higher volatility and require models capable of quick adaptation. SVR outperformed others due to its robustness in high-dimensional, noisy datasets. The study also explores feature engineering techniques like rolling averages and market sentiment scores to improve accuracy.

12. Rao et al. (2023)
Title: *AI-ML driven predictive analytics for forecasting prices of agricultural commodities*
Summary:
Rao et al. present an end-to-end predictive analytics system using Python and Streamlit for real-time forecasting. The model integrates Random Forest, SVR, and LSTM for performance benchmarking. A key innovation is the serialized Pickle model that allows fast deployment on mobile and web applications. This system is designed with farmers and agri-businesses in mind, offering actionable insights in a user-friendly format.

PROPOSED SYSTEM

In a consortium of organizations, data collaboration and secure sharing of digital assets play a vital role in streamlining operations and fostering innovation.

Forecasting agricultural commodity prices plays a pivotal role in reducing uncertainties faced by farmers, traders, policymakers, and stakeholders. Effective price prediction models allow stakeholders to make better-informed decisions on investments, production, and trade. To achieve this, a detailed and structured approach to **data collection**, **data preprocessing**, **exploratory data analysis (EDA)**, **model selection**, **model training**, and **evaluation** is essential. This framework can ensure that the models used for predicting prices are both accurate and reliable. In this section, we will expand on each of these steps, elaborating on the critical tasks involved.

1. Data Collection

The first step in the process of price forecasting is **data collection**. Agricultural prices are influenced by a variety of factors, including historical price trends, environmental conditions, and economic indicators. Therefore, comprehensive data from different sources is required to build an effective predictive model.

Sources:

- **Historical Price Data**: This data is essential for any price prediction model. Historical price data provides a time series of past prices, enabling the model to learn trends and patterns. Sources for such data include government agricultural departments, commodity exchanges, and industry reports.
- **Weather Conditions**: Weather is one of the most significant factors that influence agricultural production. Temperature, rainfall, and seasonal changes can affect both the yield and quality of crops, subsequently influencing their prices. Weather data can be sourced from meteorological agencies or global weather databases.
- **Economic Indicators**: Economic conditions such as inflation rates, exchange rates, and interest rates play a role in price determination. These economic indicators often influence the demand and supply chain. For instance, rising inflation can lead to higher commodity prices due to increased production costs.
- **Trade Data**: Trade data involves the import and export statistics of agricultural commodities, providing insights into global supply and demand. This data can help predict price fluctuations based on international market trends.

Data Types:

- **Time Series Data**: The most crucial type of data for price prediction is time series data, where the values (prices, in this case) are collected at regular time intervals. Time series data is critical for capturing trends, cycles, and seasonality, which are fundamental in forecasting agricultural commodity prices.
- **Market Sentiment Analysis**: Social media platforms, news articles, and market reports often influence the perception of commodity prices. Sentiment analysis, powered by Natural Language Processing (NLP), can gauge the overall sentiment surrounding a commodity, helping to predict price movements.
- **Environmental Factors**: External factors, such as geopolitical events, natural disasters, and technological advancements, can disrupt agricultural production and trade. By analyzing these factors, stakeholders can gain a better understanding of potential risks that might affect commodity prices.

2. Data Preprocessing

Once data has been collected, it requires **preprocessing** to ensure its quality and consistency for model development. Raw data often contains missing values, outliers, or other issues that could impair the performance of machine learning models.

Cleaning:

- **Handling Missing Values**: Agricultural data can often be incomplete, especially in emerging markets. Common strategies for handling missing values include:
    - **Imputation**: Replacing missing values with estimates, such as the mean, median, or the result of more sophisticated imputation techniques like KNN (K-Nearest Neighbors) imputation.

  - **Forward/Backward Fill**: Filling missing values by carrying forward or backward the last known data point, especially useful in time series data.
- **Removing Outliers**: Extreme values, if not reflective of actual events, can skew predictions. Detecting and removing outliers can be done using statistical techniques like Z-scores or Interquartile Range (IQR) filtering.
- **Normalizing Data**: To avoid issues with variable scales, data normalization or standardization is performed. For instance, commodity prices could range from low values to high values, while features like temperature or rainfall may be on entirely different scales. Common methods include Min-Max normalization and Z-score standardization.

Feature Engineering:

Feature engineering refers to the process of creating new variables or modifying existing ones to improve model performance.

- **Lag Features**: In time series forecasting, past values are predictive of future prices. Lag features involve creating new variables that represent the value of a feature at previous time steps (e.g., "price_last_month" or "rainfall_last_week").
- **Rolling Averages**: To capture trends in the data and smooth out short-term fluctuations, rolling averages (e.g., 7-day, 30-day) are computed for variables like commodity prices, temperature, or rainfall.
- **Seasonality Indicators**: Given that agricultural prices often exhibit seasonality (e.g., higher prices during droughts or harvest seasons), features indicating the season of the year or the crop cycle can be added to the dataset.
- **Economic Factors Integration**: It’s also essential to integrate economic variables such as inflation rates, GDP, and trade tariffs, as they can provide crucial insights into price volatility.

3. Exploratory Data Analysis (EDA)

After preprocessing, **Exploratory Data Analysis (EDA)** allows us to understand the data better and uncover underlying patterns or trends.

Trend Analysis:

Trend analysis involves examining the historical price data to identify long-term movements in the price of agricultural commodities. This is typically done by plotting the data over time to visualize any increasing or decreasing trends, cyclical patterns, or sudden shifts in prices due to events like natural disasters or market crashes.

Stationarity Checks:

A critical assumption in time series modeling is that the data should be stationary. A stationary time series has constant mean, variance, and autocovariance over time. Non-stationary data may lead to inaccurate model predictions. Statistical tests, such as the Augmented Dickey-Fuller (ADF) test, can be applied to check if the data is stationary. If the data is non-stationary, transformation methods like differencing or log transformations may be required.

Correlation Analysis:

Understanding the relationships between various variables is key to feature selection and model interpretation. For instance, price changes might be strongly correlated with rainfall in a specific region, while other variables like trade volumes might have less impact. Correlation matrices and heatmaps can visually display the degree of relationships between different features and target variables.

4. Model Selection

After the data is prepared, the next step is to select the right machine learning or deep learning model for predicting agricultural commodity prices. The choice of model depends on the complexity of the data, the amount of data available, and the desired accuracy.

Machine Learning Algorithms:

- **Random Forest**: This ensemble method works by constructing multiple decision trees and averaging their predictions. It’s especially effective for capturing non-linear relationships and can handle high-dimensional data with ease.
- **Gradient Boosting**: Another powerful ensemble learning method, Gradient Boosting builds trees sequentially, where each tree corrects the errors made by the previous ones. It often outperforms Random Forest in terms of accuracy but is more prone to overfitting.
- **Support Vector Machines (SVM)**: SVM is effective for classification and regression problems with non-linear relationships. It works well when the number of features is large compared to the number of observations, making it suitable for certain agricultural datasets.

Deep Learning Approaches:

- **Long Short-Term Memory (LSTM)**: LSTM is a type of Recurrent Neural Network (RNN) designed to handle sequences of data, making it ideal for time-series forecasting. LSTM can capture long-term dependencies in the data, making it particularly useful for price prediction in agriculture, where past data can influence future prices for extended periods.

5. Model Training & Hyperparameter Tuning

Model training is a crucial step, where the model learns from the data. Training involves splitting the data into **training** and **test sets**. The training set is used to fit the model, while the test set is used to evaluate its generalizability.

Cross-Validation:

In time-series data, traditional cross-validation (like K-fold) is not appropriate because it could lead to data leakage. **Time-series cross-validation** involves splitting the data into training and validation sets while ensuring that the validation set is always later in time than the training set. This way, we simulate the real-world forecasting scenario.

Parameter Optimization:

Hyperparameter tuning is necessary to find the best configuration for the chosen model. Techniques such as **Grid Search** (testing all possible hyperparameter combinations) and **Random Search** (randomly selecting hyperparameters from a predefined range) can be applied to find the optimal values that minimize prediction errors.

6. Model Evaluation

Finally, once a model is trained, its performance must be evaluated using appropriate metrics.

Performance Metrics:

- **Mean Absolute Error (MAE)**: This metric gives the average magnitude of the errors in predictions, without considering their direction. It's useful for assessing the overall performance.
- **Root Mean Squared Error (RMSE)**: RMSE penalizes larger errors more heavily, providing a sense of how well the model performs on large deviations in price predictions.
- **R-squared (R^2) Score**: This metric indicates how well the model's predictions match the observed data. A higher R^2 value indicates a better fit.

Residual Analysis:

Residual analysis involves plotting the residuals (the differences between predicted and actual values) to check if the model is well-calibrated. Ideally, residuals should exhibit no patterns, confirming that the model has captured all the useful information from the data.

RESULTS AND DISCUSSION

In the context of agricultural commodity price prediction, model evaluation is a critical step to assess the accuracy and robustness of forecasting models. These models, which may include machine learning and deep learning algorithms, are designed to predict commodity prices based on historical data, economic indicators, and other relevant factors. However, the success of these models hinges on their ability to make accurate predictions. To achieve this, the performance of these models needs to be measured rigorously using several evaluation metrics. The most commonly used metrics in this regard are **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and **R-squared (R^2)**. These metrics help assess how well a model fits the data and predict future prices.

In the agricultural sector, where price volatility is often influenced by various external factors such as weather conditions, geopolitical events, and market sentiment, accurate forecasting is essential. Model evaluation metrics allow stakeholders, including farmers, traders, and policymakers, to make informed decisions based on the predictive model's performance. Let's explore each of these evaluation metrics in detail:

1. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is one of the simplest and most commonly used metrics for evaluating regression models. MAE calculates the average of the absolute differences between predicted and actual values. It is a straightforward way to understand the magnitude of prediction errors and is particularly useful when the goal is to minimize the overall deviation between predicted and observed values.

Interpretation of MAE:

- **Low MAE**: A lower MAE indicates that the model's predictions are close to the actual values, meaning the model is performing well in terms of accuracy.

- **High MAE**: A higher MAE suggests that the model is making larger errors in its predictions, which can be problematic, especially in forecasting agricultural prices where precise predictions are needed for decision-making.

Since MAE treats all errors equally (by taking the absolute value of the difference), it does not penalize larger errors more than smaller ones. This is advantageous in situations where it is important to avoid large discrepancies in predictions. For instance, when predicting agricultural commodity prices, a small error in price prediction is less detrimental than a large one. MAE, therefore, is a good general-purpose metric but may not always highlight models' performance on large deviations as strongly as other metrics.

Advantages of MAE:

- Easy to interpret.
- Provides a direct measure of average model error.
- Does not require assumptions about the distribution of errors.

Limitations of MAE:

- MAE does not differentiate between large and small errors, which can be a limitation when some deviations in predictions have a more significant impact than others.

2. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is another widely used metric for evaluating regression models. RMSE takes the average of the squared differences between predicted and actual values, then takes the square root of that average. The primary advantage of RMSE over MAE is that it penalizes larger errors more heavily due to the squaring of differences. This characteristic makes RMSE more sensitive to large errors, which can be useful when large prediction deviations are particularly undesirable.

- **Low RMSE**: A lower RMSE indicates that the model performs well in terms of both accuracy and consistency. The closer the predicted values are to the actual values, the lower the RMSE will be.
- **High RMSE**: A higher RMSE suggests that the model is consistently making larger errors, which can significantly affect the quality of the model's predictions.

Unlike MAE, RMSE gives more importance to large errors, meaning it is particularly useful when large deviations from the true price are undesirable in agricultural price prediction. For instance, an overestimate of commodity prices by 10% could lead to wasted investments, while an underestimate of 10% could result in lost profit opportunities. Thus, RMSE ensures that models minimize the chances of making significant prediction errors.

Advantages of RMSE:

- Penalizes larger errors, making it more sensitive to outliers.
- Provides a better indication of how well the model fits the data when large errors are especially problematic.
- Suitable for problems where large deviations have a significant impact on outcomes, such as in pricing decisions.

Limitations of RMSE:

- Since RMSE squares the errors, it is more sensitive to outliers, which might lead to the model being overly penalized for a few large errors.
- The interpretation of RMSE can be more difficult than MAE because the error term is in the squared units of the predicted variable (e.g., squared currency units).

3. R-squared (R^2)

R-squared (R^2) is a metric that evaluates how well the independent variables (or features) in a model explain the variability in the dependent variable (or target). In simple terms, R^2 measures the proportion of the variance in the observed data that is explained by the model. An R^2 value close to 1 indicates that the model explains a high proportion of the variance, while a value close to 0 indicates poor explanatory power.

Interpretation of R^2:

- **$R^2 = 1$**: The model perfectly predicts the target variable, and there is no unexplained variance.
- **$R^2 = 0$**: The model does no better than simply predicting the mean value of the target variable for all observations.
- **Negative R^2**: In some cases, R^2 can be negative if the model performs worse than simply predicting the mean value, indicating poor model performance.

R^2 is a useful metric for understanding how well the model explains the data. In the case of agricultural price forecasting, if the R^2 value is high, it indicates that the model can effectively capture the trends and variations in commodity prices, which is essential for accurate decision-making. Conversely, a low R^2 suggests that the model is unable to capture important price movements, thus limiting its usefulness.

Advantages of R^2:

- Easy to understand and interpret as it represents the proportion of variance explained by the model.
- Useful for comparing models with different levels of complexity.

Limitations of R^2:

- R^2 can be misleading if the model has too many variables (overfitting), as it may give a false sense of predictive power.
- It is not a good metric when comparing models with different types of data (e.g., time series data versus static data).

CONCLUSION

In conclusion, forecasting agricultural commodity prices is a highly complex yet essential task for farmers, traders, policymakers, and other stakeholders who rely on accurate price predictions for effective decision-making. This process requires sophisticated machine learning models that can integrate multiple factors, such as historical price data, weather conditions, economic indicators, and market sentiment, to predict future price trends. The evaluation of these models is crucial for ensuring their accuracy and reliability, and three primary metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2)—serve as the cornerstone of this assessment. MAE offers a straightforward measure of the average error in predictions, providing a clear indication of model accuracy, while RMSE goes a step further by penalizing larger errors, making it more sensitive to significant deviations in predicted values, which is vital when large price fluctuations can have profound economic impacts. R^2, on the other hand, assesses how well the model explains the variance in the observed data, offering valuable insight into its overall explanatory power. By employing these metrics, we gain a comprehensive understanding of how well the models are performing, and can identify areas for improvement, whether it's reducing error magnitude or improving the model's ability to capture underlying patterns in the data. Each of these metrics has its strengths and weaknesses, and when used together, they provide a nuanced picture of model performance, guiding further refinements and optimizations. The integration of advanced machine learning techniques such as Random Forest, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks, combined with these robust evaluation metrics, holds the potential to revolutionize agricultural price forecasting. With these methods, stakeholders can better navigate the inherent uncertainties of the agricultural market, making more informed decisions that can reduce risk, enhance financial planning, and ultimately contribute to the stability and sustainability of the agricultural sector. Therefore, it is clear that continuous refinement of forecasting models, driven by rigorous evaluation and the use of sophisticated machine learning algorithms, is essential to improving the accuracy and reliability of agricultural commodity price predictions, enabling more effective and efficient decision-making across the agricultural value chain.

REFERENCES

1. Reddy, C. N. K., & Murthy, G. V. (2012). Evaluation of Behavioral Security in Cloud Computing. *International Journal of Computer Science and Information Technologies*, *3*(2), 3328-3333.
2. Murthy, G. V., Kumar, C. P., & Kumar, V. V. (2017, December). Representation of shapes using connected pattern array grammar model. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 819-822). IEEE.
3. Krishna, K. V., Rao, M. V., & Murthy, G. V. (2017). Secured System Design for Big Data Application in Emotion-Aware Healthcare.
4. Rani, G. A., Krishna, V. R., & Murthy, G. V. (2017). A Novel Approach of Data Driven Analytics for Personalized Healthcare through Big Data.
5. Rao, M. V., Raju, K. S., Murthy, G. V., & Rani, B. K. (2020). Configure and Management of Internet of Things. *Data Engineering and Communication Technology*, 163.
6. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
7. Chithanuru, V., & Ramaiah, M. (2023). An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions–A review. *Concurrency and Computation: Practice and Experience*, *35*(22), e7724.
8. Prashanth, J. S., & Nandury, S. V. (2015, June). Cluster-based rendezvous points selection for reducing tour length of mobile element in WSN. In *2015 IEEE International Advance Computing Conference (IACC)* (pp. 1230-1235). IEEE.
9. Kumar, K. A., Pabboju, S., & Desai, N. M. S. (2014). Advance text steganography algorithms: an overview. *International Journal of Research and Applications*, *1*(1), 31-35.
10. Hnamte, V., & Balram, G. (2022). Implementation of Naive Bayes Classifier for Reducing DDoS Attacks in IoT Networks. *Journal of Algebraic Statistics*, *13*(2), 2749-2757.
11. Balram, G., Anitha, S., & Deshmukh, A. (2020, December). Utilization of renewable energy sources in generation and distribution optimization. In *IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 4, p. 042054). IOP Publishing.
12. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
13. Mahammad, F. S., Viswanatham, V. M., Tahseen, A., Devi, M. S., & Kumar, M. A. (2024, July). Key distribution scheme for preventing key reinstallation attack in wireless networks. In *AIP Conference Proceedings* (Vol. 3028, No. 1). AIP Publishing.
14. Lavanya, P. (2024). In-Cab Smart Guidance and support system for Dragline operator.
15. Kovoor, M., Durairaj, M., Karyakarte, M. S., Hussain, M. Z., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. *Measurement: Sensors*, *32*, 101054.
16. Rao, N. R., Kovoor, M., Kishor Kumar, G. N., & Parameswari, D. V. L. (2023). Security and privacy in smart farming: challenges and opportunities. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(7).
17. Madhuri, K. (2023). Security Threats and Detection Mechanisms in Machine Learning. *Handbook of Artificial Intelligence*, *255*.
18. Reddy, B. A., & Reddy, P. R. S. (2012). Effective data distribution techniques for multi-cloud storage in cloud computing. *CSE, Anurag Group of Institutions, Hyderabad, AP, India*.

19. Srilatha, P., Murthy, G. V., & Reddy, P. R. S. (2020). Integration of Assessment and Learning Platform in a Traditional Class Room Based Programming Course. *Journal of Engineering Education Transformations*, *33*, 179-184.
20. Reddy, P. R. S., & Ravindranadh, K. (2019). An exploration on privacy concerned secured data sharing techniques in cloud. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1190-1198.
21. Raj, R. S., & Raju, G. P. (2014, December). An approach for optimization of resource management in Hadoop. In *International Conference on Computing and Communication Technologies* (pp. 1-5). IEEE.
22. Ramana, A. V., Bhoga, U., Dhulipalla, R. K., Kiran, A., Chary, B. D., & Reddy, P. C. S. (2023, June). Abnormal Behavior Prediction in Elderly Persons Using Deep Learning. In *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)* (pp. 1-5). IEEE.
23. Yakoob, S., Krishna Reddy, V., & Dastagiraiah, C. (2017). Multi User Authentication in Reliable Data Storage in Cloud. In *Computer Communication, Networking and Internet Security: Proceedings of IC3T 2016* (pp. 531-539). Springer Singapore.
24. Sukhavasi, V., Kulkarni, S., Raghavendran, V., Dastagiraiah, C., Apat, S. K., & Reddy, P. C. S. (2024). Malignancy Detection in Lung and Colon Histopathology Images by Transfer Learning with Class Selective Image Processing.
25. Dastagiraiah, C., Krishna Reddy, V., & Pandurangarao, K. V. (2018). Dynamic load balancing environment in cloud computing based on VM ware off-loading. In *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016* (pp. 483-492). Springer Singapore.
26. Swapna, N. (2017). „Analysis of Machine Learning Algorithms to Protect from Phishing in Web Data Mining‟. *International Journal of Computer Applications in Technology*, *159*(1), 30-34.
27. Moparthi, N. R., Bhattacharyya, D., Balakrishna, G., & Prashanth, J. S. (2021). Paddy leaf disease detection using CNN.
28. Balakrishna, G., & Babu, C. S. (2013). Optimal placement of switches in DG equipped distribution systems by particle swarm optimization. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(12), 6234-6240.
29. Moparthi, N. R., Sagar, P. V., & Balakrishna, G. (2020, July). Usage for inside design by AR and VR technology. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.
30. Amarnadh, V., & Moparthi, N. R. (2023). Comprehensive review of different artificial intelligence-based methods for credit risk assessment in data science. *Intelligent Decision Technologies*, *17*(4), 1265-1282.
31. Amarnadh, V., & Moparthi, N. (2023). Data Science in Banking Sector: Comprehensive Review of Advanced Learning Methods for Credit Risk Assessment. *International Journal of Computing and Digital Systems*, *14*(1), 1-xx.
32. Amarnadh, V., & Rao, M. N. (2025). A Consensus Blockchain-Based Credit Risk Evaluation and Credit Data Storage Using Novel Deep Learning Approach. *Computational Economics*, 1-34.
33. Shailaja, K., & Anuradha, B. (2017). Improved face recognition using a modified PSO based self-weighted linear collaborative discriminant regression classification. *J. Eng. Appl. Sci*, *12*, 7234-7241.
34. Sekhar, P. R., & Goud, S. (2024). Collaborative Learning Techniques in Python Programming: A Case Study with CSE Students at Anurag University. *Journal of Engineering Education Transformations*, *38*.
35. Sekhar, P. R., & Sujatha, B. (2023). Feature extraction and independent subset generation using genetic algorithm for improved classification. *Int. J. Intell. Syst. Appl. Eng*, *11*, 503-512.
36. Pesaramelli, R. S., & Sujatha, B. (2024, March). Principle correlated feature extraction using differential evolution for improved classification. In *AIP Conference Proceedings* (Vol. 2919, No. 1). AIP Publishing.
37. Tejaswi, S., Sivaprashanth, J., Bala Krishna, G., Sridevi, M., & Rawat, S. S. (2023, December). Smart Dustbin Using IoT. In *International Conference on Advances in Computational Intelligence and Informatics* (pp. 257-265). Singapore: Springer Nature Singapore.
38. Moreb, M., Mohammed, T. A., & Bayat, O. (2020). A novel software engineering approach toward using machine learning for improving the efficiency of health systems. *IEEE Access*, *8*, 23169-23178.
39. Ravi, P., Haritha, D., & Niranjan, P. (2018). A Survey: Computing Iceberg Queries. *International Journal of Engineering & Technology*, *7*(2.7), 791-793.
40. Madar, B., Kumar, G. K., & Ramakrishna, C. (2017). Captcha breaking using segmentation and morphological operations. *International Journal of Computer Applications*, *166*(4), 34-38.
41. Rani, M. S., & Geetavani, B. (2017, May). Design and analysis for improving reliability and accuracy

of big-data based peripheral control through IoT. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 749-753). IEEE.

42. Reddy, T., Prasad, T. S. D., Swetha, S., Nirmala, G., & Ram, P. (2018). A study on antiplatelets and anticoagulants utilisation in a tertiary care hospital. *International Journal of Pharmaceutical and Clinical Research*, *10*, 155-161.
43. Prasad, P. S., & Rao, S. K. M. (2017). HIASA: Hybrid improved artificial bee colony and simulated annealing based attack detection algorithm in mobile ad-hoc networks (MANETs). *Bonfring International Journal of Industrial Engineering and Management Science*, *7*(2), 01-12.
44. AC, R., Chowdary Kakarla, P., Simha PJ, V., & Mohan, N. (2022). Implementation of Tiny Machine Learning Models on Arduino 33–BLE for Gesture and Speech Recognition.
45. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
46. Nagaraj, P., Prasad, A. K., Narsimha, V. B., & Sujatha, B. (2022). Swine flu detection and location using machine learning techniques and GIS. *International Journal of Advanced Computer Science and Applications*, *13*(9).
47. Priyanka, J. H., & Parveen, N. (2024). DeepSkillNER: an automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach. *Multimedia Tools and Applications*, *83*(16), 47503-47530.
48. Sathish, S., Thangavel, K., & Boopathi, S. (2010). Performance analysis of DSR, AODV, FSR and ZRP routing protocols in MANET. *MES Journal of Technology and Management*, 57-61.
49. Siva Prasad, B. V. V., Mandapati, S., Kumar Ramasamy, L., Boddu, R., Reddy, P., & Suresh Kumar, B. (2023). Ensemble-based cryptography for soldiers' health monitoring using mobile ad hoc networks. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *64*(3), 658-671.
50. Elechi, P., & Onu, K. E. (2022). Unmanned Aerial Vehicle Cellular Communication Operating in Non-terrestrial Networks. In *Unmanned Aerial Vehicle Cellular Communications* (pp. 225-251). Cham: Springer International Publishing.
51. Prasad, B. V. V. S., Mandapati, S., Haritha, B., & Begum, M. J. (2020, August). Enhanced Security for the authentication of Digital Signature from the key generated by the CSTRNG method. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1088-1093). IEEE.
52. Mukiri, R. R., Kumar, B. S., & Prasad, B. V. V. (2019, February). Effective Data Collaborative Strain Using RecTree Algorithm. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*.
53. Balaraju, J., Raj, M. G., & Murthy, C. S. (2019). Fuzzy-FMEA risk evaluation approach for LHD machine–A case study. *Journal of Sustainable Mining*, *18*(4), 257-268.
54. Thirumoorthi, P., Deepika, S., & Yadaiah, N. (2014, March). Solar energy based dynamic sag compensator. In *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)* (pp. 1-6). IEEE.
55. Vinayasree, P., & Reddy, A. M. (2025). A Reliable and Secure Permissioned Blockchain-Assisted Data Transfer Mechanism in Healthcare-Based Cyber-Physical Systems. *Concurrency and Computation: Practice and Experience*, *37*(3), e8378.
56. Acharjee, P. B., Kumar, M., Krishna, G., Raminenei, K., Ibrahim, R. K., & Alazzam, M. B. (2023, May). Securing International Law Against Cyber Attacks through Blockchain Integration. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2676-2681). IEEE.
57. Ramineni, K., Reddy, L. K. K., Ramana, T. V., & Rajesh, V. (2023, July). Classification of Skin Cancer Using Integrated Methodology. In *International Conference on Data Science and Applications* (pp. 105-118). Singapore: Springer Nature Singapore.
58. LAASSIRI, J., EL HAJJI, S. A. Ï. D., BOUHDADI, M., AOUDE, M. A., JAGADISH, H. P., LOHIT, M. K., ... & KHOLLADI, M. (2010). Specifying Behavioral Concepts by engineering language of RM-ODP. *Journal of Theoretical and Applied Information Technology*, *15*(1).
59. Prasad, D. V. R., & Mohanji, Y. K. V. (2021). FACE RECOGNITION-BASED LECTURE ATTENDANCE SYSTEM: A SURVEY PAPER. *Elementary Education Online*, *20*(4), 1245-1245.
60. Dasu, V. R. P., & Gujjari, B. (2015). Technology-Enhanced Learning Through ICT Tools Using Aakash Tablet. In *Proceedings of the International Conference on Transformations in Engineering Education: ICTIEE 2014* (pp. 203-216). Springer India.
61. Reddy, A. M., Reddy, K. S., Jayaram, M., Venkata Maha Lakshmi, N., Aluvalu, R., Mahesh, T. R., ...

& Stalin Alex, D. (2022). An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network. *Journal of Sensors*, *2022*(1), 4093658.

62. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S., Hasane Ahammad, S., Kumar, G. N. S., & Mallikarjuna Reddy, A. (2019). Texture filtration module under stabilization via random forest optimization methodology. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(3), 458-469.
63. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
64. Sirisha, G., & Reddy, A. M. (2018, September). Smart healthcare analysis and therapy for voice disorder using cloud and edge computing. In *2018 4th international conference on applied and theoretical computing and communication technology (iCATccT)* (pp. 103-106). IEEE.
65. Reddy, A. M., Yarlagadda, S., & Akkinen, H. (2021). An extensive analytical approach on human resources using random forest algorithm. *arXiv preprint arXiv:2105.07855*.
66. Kumar, G. N., Bhavanam, S. N., & Midasala, V. (2014). Image Hiding in a Video-based on DWT & LSB Algorithm. In *ICPVS Conference*.
67. Naveen Kumar, G. S., & Reddy, V. S. K. (2022). High performance algorithm for content-based video retrieval using multiple features. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021* (pp. 637-646). Singapore: Springer Nature Singapore.
68. Reddy, P. S., Kumar, G. N., Ritish, B., SaiSwetha, C., & Abhilash, K. B. (2013). Intelligent parking space detection system based on image segmentation. *Int J Sci Res Dev*, *1*(6), 1310-1312.
69. Naveen Kumar, G. S., Reddy, V. S. K., & Kumar, S. S. (2018). High-performance video retrieval based on spatio-temporal features. *Microelectronics, Electromagnetics and Telecommunications*, 433-441.
70. Kumar, G. N., & Reddy, M. A. BWT & LSB algorithm based hiding an image into a video. *IJESAT*, 170-174.
71. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
72. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
73. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024, August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.
74. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
75. Anuprathibha, T., Praveen, R. V. S., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using Text and Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.
76. Shinkar, A. R., Joshi, D., Praveen, R. V. S., Rajesh, Y., & Singh, D. (2024, December). Intelligent solar energy harvesting and management in IoT nodes using deep self-organizing maps. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.
77. Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). AI Powered Plant Identification and Plant Disease Classification System. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1610-1616). IEEE.
78. Dhivya, R., Sagili, S. R., Praveen, R. V. S., VamsiLala, P. N. V., Sangeetha, A., & Suchithra, B. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 997-1002). IEEE.
79. Kemmannu, P. K., Praveen, R. V. S., Saravanan, B., Amshavalli, M., & Banupriya, V. (2024, December). Enhancing Sustainable Agriculture Through Smart Architecture: An Adaptive Neuro-Fuzzy Inference System with XGBoost Model. In *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)* (pp. 724-730). IEEE.

80. Praveen, R. V. S. (2024). *Data Engineering for Modern Applications*. Addition Publishing House.