Prediction of Agricultural Commodity Prices using Big Data Framework

Humaira Rana
Department of Computer Science & Information Technology, NED University of Engineering and Technology, Pakistan
engr.humairarana@gmail.com (corresponding author)

Muhammad Umer Farooq
Department of Computer Science & Information Technology, NED University of Engineering and Technology, Pakistan
umer@neduet.edu.pk

Abdul Karim Kazi
Department of Computer Science & Information Technology, NED University of Engineering and Technology, Pakistan
karimkazi@neduet.edu.pk

Mirza Adnan Baig
Department of Computer Science, IQRA University, Pakistan
adnanbaig996@gmail.com

Muhammad Ali Akhtar
Department of Computer and Information Systems Engineering, NED University of Engineering and Technology, Pakistan
aliakhtar@neduet.edu.pk

Received: 4 October 2023 | Revised: 30 October 2023 | Accepted: 4 November 2023

 DOI: https://doi.org/10.48084/etasr.6468

ABSTRACT

The agriculture sector plays a crucial role in the economy of Pakistan, contributing significantly to the Gross Domestic Product (GDP) and the employment rate. However, this sector faces challenges such as climate change, water scarcity, and low productivity, which have a direct impact on agricultural commodity prices. Accurate forecasting of commodity prices is essential for farmers, traders, and policymakers to make informed decisions and improve economic outcomes. This paper explores the use of a big data framework for agricultural commodity price forecasting in Pakistan, using a historical dataset on commodity prices in various Pakistani cities from 2007 to 2022 and Apache Spark to preprocess and clean the data. Based on historical spinach prices in Vehari City, the machine learning models Auto-Regressive Moving Average (ARIMA), Random Forest, and Long-Short-Term Memory (LSTM) were applied to price trends, and their performance was compared using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and squared correlation coefficient (R^2). LSTM outperformed ARIMA and Random Forest with a higher R^2 value of 0.8 and the lowest MAE of 125.29. Such predictions can help farmers to effectively plan crop cultivation and traders to make well-informed decisions.

Keywords-agricultural commodity; price forecasting; big data analytics; Apache Spark framework; Pyspark

I. INTRODUCTION

Agricultural commodities play an essential role in our daily lives and have a significant influence on economic stability. Fluctuations in the prices of these commodities can impose a burden on consumers and disrupt the income stability of farm households. Furthermore, abnormal climate patterns observed in recent years have exacerbated the volatility of agricultural commodity prices, requiring governments to shape effective policies and decisions to maintain supply-demand equilibrium

[1]. Over the past year, rising food prices, coupled with rising energy costs, have eroded the purchasing power of consumers around the world. However, this impact is not uniform across all countries. The International Monetary Fund highlights that low-income countries allocate a higher proportion of their consumer spending to food products compared to high-income countries. In these lower-income countries, food expenditure can soar up to 44% of disposable income, while it averages around 28% in emerging countries and drops to 16% in developed economies. This disparity has resulted in a notable increase in inflation rates, particularly in lower-income and certain emerging countries, where many of them experience double-digit inflation [2].

The agriculture sector is a vital component of the Pakistani economy, contributing significantly to the country's Gross Domestic Product (GDP) and employment rate. According to the Pakistan Bureau of Statistics, agriculture accounts for around 19.2% of the country's GDP and employs almost 38.5% of the workforce [3]. Pakistan has been estimated to possess around 30,930,000 hectares of agricultural land, almost 47.09% of the total land area [4]. The proportion of agricultural land and population is 0.186 ha/person, which is rather low compared to European countries [5]. This sector also faces several challenges, including climate change, water scarcity, and low productivity levels. These challenges, combined with other factors, such as global economic conditions and government policies, have a significant impact on agricultural commodity prices in Pakistan [6]. Fluctuations in commodity prices in Pakistan have a direct and far-reaching impact on the country's economy, particularly on the rural population. Farmers, who constitute a significant proportion of the rural population, are heavily dependent on agricultural commodity prices for their livelihood. For instance, a drop in cotton prices can negatively affect the income of cotton growers, thereby leading to social and economic implications for their families and communities.

The pricing of future market commodities is a matter of great importance for governments, investors, and producers alike. Food commodities, in particular, are traded in localized markets where supply and pricing are closely linked. The impact of commodities on local economies and food security for the poor is significant, and their prices are highly susceptible to fluctuations caused by multiple factors, such as investments, production, global financial outlook, and monetary policies [7]. Given these complex and dynamic factors, reliable commodity price forecasting is crucial for investors and governments to minimize the risks associated with price volatility. Forecasting is based primarily on past and current knowledge and is used to predict production and assess the parameters that will influence production in the coming years. This is done to encourage people to grow the most profitable crops and to attract non-farmers into start farming. In Pakistan, the agriculture department is responsible for forecasting the major crop production, mainly to estimate supply and demand [8]. Accurate forecasts will help farmers make informed decisions about what crops to plant and when to sell their production. They will also help traders and policymakers plan their trading and policy decisions, leading to more stable prices and better economic outcomes.

With the advent of big data technologies and advanced analytics tools, it is now possible to take advantage of the vast amount of data generated in the agricultural sector and use it to predict commodity prices with greater accuracy [9-10]. Apache Spark is one such technology that has gained popularity in recent years and is an open-source big data processing framework that enables scalable and distributed data processing [11]. Machine learning, a subdomain of artificial intelligence, provides powerful algorithms that can be trained to perform specific tasks or achieve specific goals. These algorithms have shown great promise in improving the accuracy of commodity price forecasting [12]. This study aims to contribute to understanding the use of big data in forecasting agricultural commodity prices in Pakistan. The results would be valuable to farmers, traders, and policymakers in making informed decisions and improving their economic outcomes. This paper explores the use of a big data framework for forecasting agricultural commodity prices using the Pyspark library. To achieve these objectives, the study uses a dataset of historical commodity prices of various cities in Pakistan. The data were preprocessed and cleaned using Apache Spark, and machine learning models Auto-Regressive Moving Average (ARIMA), Random Forest, and Long Short Term Memory (LSTM). The performance of these models was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 values.

In [13], an econometric analysis was used to explore the relationship between Agricultural GDP (AGDP) and fruits in Pakistan. The Augmented Dickey-Fuller (ADF) and the Ordinary Least Square (OLS) tests were used for data analysis, while the Johansen co-integration test was used to interpret the results, showing the negative and insignificant impact of apples and grapes on AGDP, and the positive and significant influence of bananas, citrus, and pears. A linear regression method was used to predict fruit production. In [14], a model for oil palm yield prediction was proposed, incorporating 420 months of yield records and 12 weather and 3 soil moisture parameters. An automated model selection process was used and the best regression models were found to be AdaBoost and Extra Tree, proving that rain days count, root zone soil wetness, wind speed, cloud amount, and rainfall were the most influential factors and have a significant impact on oil palm yield [14]. In [15], garlic prices were predicted using an ARIMA-SVM method with an RMSE of 1.1131. The ARIMA model was used for the linear part prediction, while the Support Vector Machine (SVM) was used for the nonlinear prediction. In [16], crop price prediction was carried out using the Random Forest Ensemble Learning (RFEL) and the Decision Tree Regressor (DTR) model. RFEL generated good results with 97.57% accuracy, while the accuracy of DT was 92.66%. In [17], a portal was developed for growers to watch the historical and future prices of selected commodities. In [18], neural networks were shown to be the best predictive models in terms of RMSE, MAPE, and MAE. In [19], an accurate wheat production forecasting model was designed with the help of Long Short Term Memory (LSTM) neural networks. LSTM outperformed ARIMA and RNN with a minimum MAE and RMSE, and the highest R value. In [20], big data were used to analyze the wholesale price of eggs, price fluctuations, and influencing factors.

Although significant studies have been conducted on the use of big data analytics, machine learning, and econometric analysis in the agricultural sector, there is still a significant gap in the proper use of these technologies and frameworks to aid Pakistani farmers in decision-making and crop planning throughout the crop cycle. The existing studies have provided valuable insights into specific aspects, such as crop yield prediction, price forecasting, and the impact of certain factors on agricultural production. However, there is a lack of comprehensive research that integrates these approaches to develop a holistic decision-making support system for farmers in Pakistan.

II. METHODOLOGY

This study encompassed a variety of essential components, including data collection and preprocessing, exploratory data analysis using Databricks software, an optimized platform for Apache Spark, the implementation of predictive modeling with distinct algorithms, and the evaluation of model performance using various metrics. These methods and tools were meticulously considered to address the multifaceted nature of agricultural price data, ensuring the robustness and effectiveness of the study.

The dataset was divided into training and validation data. Three different algorithms, namely ARIMA, Random Forest (RF), and LSTM were used to forecast crop prices. This choice was guided by the distinct strengths and applicability of these methods to the specific task of crop price forecasting. ARIMA was selected for its effectiveness in modeling time series data, capturing linear trends, and seasonality. However, it may struggle with complex, nonlinear relationships. On the other hand, RF excels at handling intricate, nonlinear patterns and is less sensitive to outliers. LSTM was chosen for its ability to capture sequential dependencies and long-term patterns in time-series data, which makes it well-suited for the task. However, it may require a larger dataset and is more complicated to configure. These model choices collectively address the diverse and dynamic nature of agricultural price data, offering a comprehensive approach for effective forecasting and model evaluation. The models were trained using the training dataset and their performance was evaluated with MSE, RMSE, MAE, and R^2. The obtained results from each model were compared to determine the most accurate forecasting method. This comparison provided insights into the performance and effectiveness of the different algorithms in predicting crop prices. Figure 1 shows the method used to ensure a comprehensive analysis of the crop price dataset and identify the most suitable forecasting approach.

A. Data Collection

Data were extracted from the Agriculture Marketing Information System (AMIS) [21], which is a marketing wing of the Punjab government's agriculture department in Pakistan, to collect, compile, and distribute agricultural market information to traders, farmers, and other relevant stakeholders in the agricultural value chain.

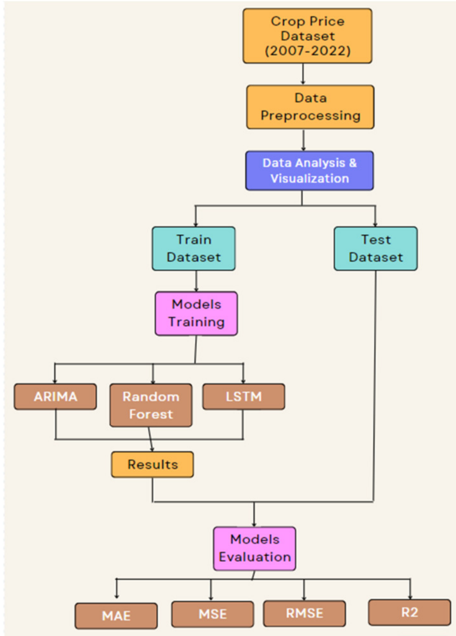

Fig. 1. Methodology.

B. Data Preprocessing

A total of 17 million values were queried for market prices of multiple crops in multiple cities of Punjab from 2007 to 2022, with features city, crop, date, month, year, and price. There were 139 unique cities and 119 unique crop varieties available in the dataset. Figure 2 shows a comprehensive visualization of the maximum prices of crops across all cities.

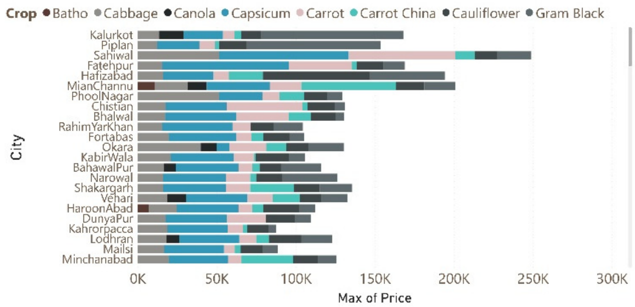

Fig. 2. Stacked bar chart showing maximum crop prices across all cities.

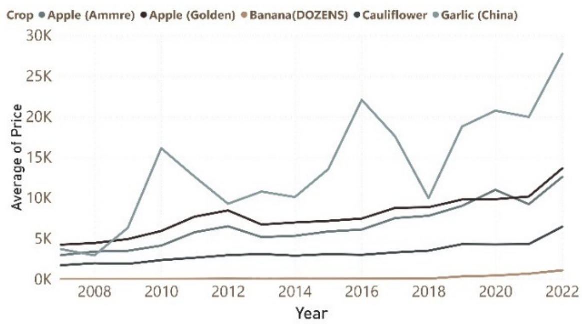

Fig. 3. Line plot illustrating the price trends of different crops.

Data cleaning and preparation tasks were performed on the dataset such as renaming columns, removing unnecessary columns and duplicates, handling missing values, filling null values with the previous non-null value using the forward fill method, removing outliers, and conversion of data types. Every combination of city and price columns was an individual time series, as shown in Figure 3.

C. Data Analysis and Machine Learning Models

Historically, spinach prices in Vehari, like other agricultural commodities, have shown some degree of volatility due to fluctuations in market conditions. During peak seasons, when the supply of spinach is abundant, prices tend to be relatively lower. Conversely, during off-peak seasons or when there are supply disruptions, prices may rise. The spinach prices in Vehari city were selected for this study and predictive tasks. Figure 4 shows a line plot visualization of the price trend of spinach in Vehari. The trend has shown quite interesting behavior. It can be observed that in 2008, 2009, and 2014, spinach prices went quite high on specific dates, and seasonality can also be observed in the trend on a yearly basis. The ADF stationarity test was applied to this time series. Table I shows the results, which examined the presence of a unit root in the data and indicated non-stationarity.

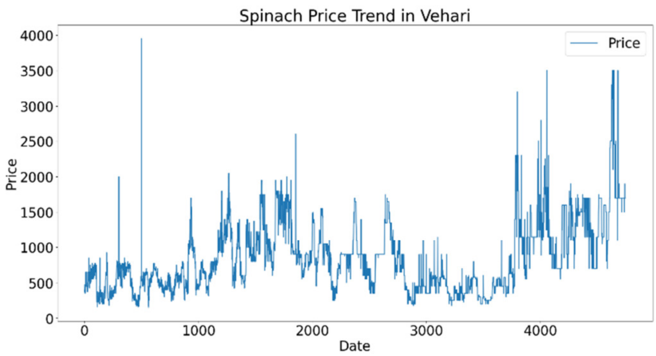

Fig. 4. Line plot depicting the price trend of spinach in Vehari.

TABLE I. ADF TEST RESULTS

ADF Test Parameters	ADF Test Values
Test Statistic	-5.506330
p-value	0.000002
Critical Value (1%)	-3.431730
Critical Value (5%)	-2.862150
Critical Value (10%)	-2.567095

The test statistic of -5.506330 indicates strong evidence against the null hypothesis of non-stationarity. The p-value of 0.000002 represents the probability of obtaining a test statistic as extreme as the observed one, assuming that the null hypothesis is true. With such a low p-value, there is strong evidence to reject the null hypothesis of non-stationarity. The critical values determined for different levels of significance (1%, 5%, and 10%) are threshold values used to compare with the test statistic. If the test statistic is more extreme than the critical value, i.e. smaller in absolute value, it provides stronger evidence against the null hypothesis. Therefore, it can be concluded that the time series is stationary.

Before applying ARIMA, it is critical to check for the presence of seasonality in the dataset. Figure 5 shows the four components of the decomposition (observed, trend, seasonal, and residual). The trend plot shows that, in general, there was an increase in crop prices from 2008 to 2013, and then a decline was observed until 2019. Subsequently, the spinach price showed an exponential increase. The seasonal graph represents the presence of a yearly seasonality in the data.

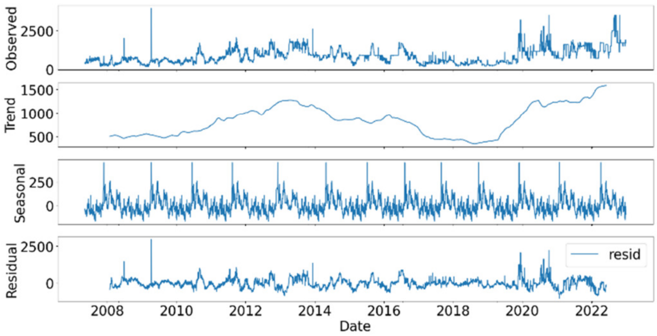

Fig. 5. Four components after decomposition.

The data were subsequently divided into training and testing sets using the random split function. The split ratio was set to 80% for training and 20% for testing data, with a specific random seed provided for reproducibility. ARIMA [22] was then applied to the dataset using the pmdarima package in Python. The auto-arima function from pmdarima.arima was used to fit an ARIMA model to the training data. This function automatically selects the optimal parameters for the ARIMA model based on the training data. The model was then used to make predictions on the testing data. The selected ARIMA model for this dataset was (1,1,1)(0,0,0) [12].

A RandomForestRegressor object was instantiated, specifying the input features of the data, i.e. Date, City and Crop, the target column of the data ("Price"), and a random seed of 42 that ensures that randomness in the model can be reproduced in the future. This object represents the RF regression model that will be trained. The model was trained by invoking the fit method on the RandomForestRegressor object and passing the training data. The resulting trained model was stored in the "model" variable. Predictions were generated on the test data using the trained model by invoking the transform method and passing the test data. Predictions were stored in the "predictions" variable.

For LSTM, the dataset was normalized in the range of 0-1 using the MinMaxScaler function. Train and test data were reshaped to fit the LSTM model. Model compile is a method in Keras, which is a high-level neural network API in Tensor Flow, used to configure the model learning process before training. It specifies the loss function and the optimizer used to minimize the loss during training. Here, the loss function was set to MSE, which is a commonly used loss function for regression problems. MSE calculates the average of the squared differences between the predicted and actual values. The Adam optimizer was used, which is also a commonly used optimizer in neural networks. It is an adaptive learning rate optimization algorithm that is designed to work well for a wide range of problems. Look-Back was used to define the number of previous time steps to consider when creating the time series dataset to train and test the LSTM model. Here, the Look-Back was set to 60, meaning that the model considered the previous 60 time steps to predict the next one. The LSTM model consists of two LSTM layers: the first LSTM layer with 50 units and a "return sequences" parameter set to True allowed it to capture sequential patterns in the input data. The second LSTM layer with 50 units followed, building on the learned sequential patterns. These LSTM layers were followed by a single dense (fully connected) layer with one unit, serving as the output layer. The neural network was trained on training data for 100 epochs and with 64 batch size. After training the model, the predictions were generated on test data using the predict function. This architecture enables the neural network to effectively capture temporal dependencies and generate accurate predictions for time-series data.

III. RESULTS AND DISCUSSION

Figure 6 shows the results of ARIMA (training data, testing data, and predictions), plotted using matplotlib. This plot reveals important insights into the model performance when applied to the test dataset. The graph displays two lines: the orange line represents the actual values obtained from the test dataset, while the green line represents the predicted values generated by the ARIMA model. However, it becomes apparent that the predicted values exhibit a straight pattern and do not closely adhere to the actual data points. This indicates that the ARIMA model may not adequately capture the underlying patterns or dynamics of the data.

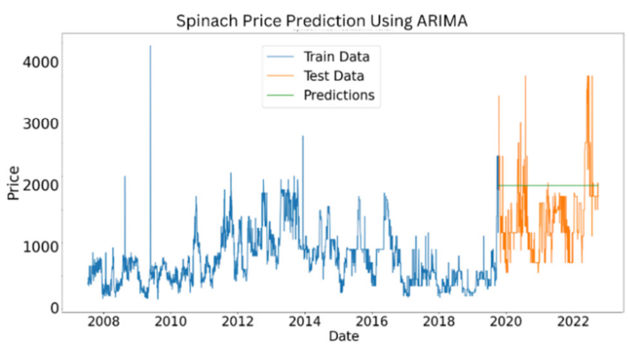

Fig. 6. Prediction results of ARIMA model.

The RF model was used to overcome the limitations of the ARIMA model, as shown in Figure 7. The decision to use the RF algorithm was based on its ability to handle complex relationships and capture nonlinear patterns within the data. The utilization of the RF model brought about notable improvements as it successfully captured the nonlinear patterns present within the data. Unlike the ARIMA model, the RF algorithm is well-suited for handling complex relationships and nonlinear dependencies. The predictions obtained from the RF model exhibited a better fit to the actual data compared to ARIMA. However, despite the model's ability to capture nonlinear patterns, the accuracy of the results was not entirely satisfactory. This result suggests that there may be additional factors or complexities at play that have not been adequately accounted for in the modeling process.

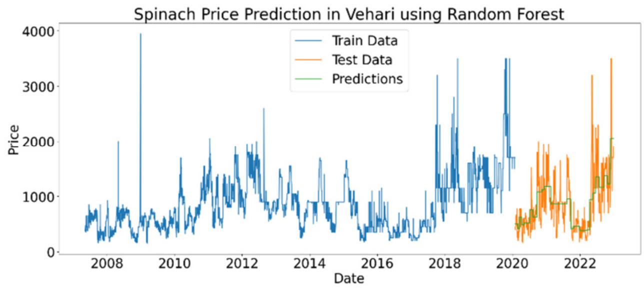

Fig. 7. Prediction results of Random Forest.

In contrast to the previous models, the LSTM model [23] demonstrated exceptional performance by perfectly replicating the test data, as shown in Figure 8. The predictions generated by the LSTM model are closely aligned with the actual values, indicating its ability to capture intricate temporal dependencies and complex patterns within the data. This result highlights the strength of LSTM models in modeling sequential data, making them particularly well-suited for time series forecasting tasks. The accurate replication of test data by the LSTM model instills confidence in its predictive capabilities and suggests that it effectively captures the underlying dynamics of the data.

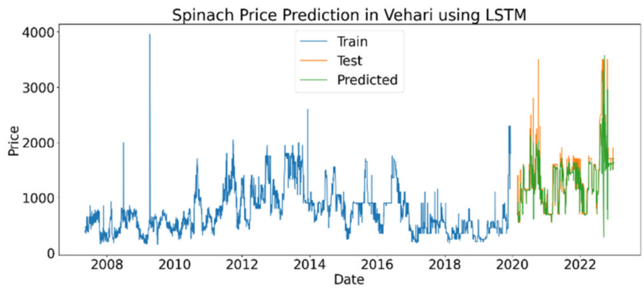

Fig. 8. Prediction results of LSTM.

The MAE, MSE, and RMSE evaluation metrics help measure the difference between the predicted and the actual values, giving an idea of how well the model performed in making predictions. Table II shows the comparison of evaluation metrics of ARIMA, RF, and LSTM. LSTM achieved the lowest MAE and MSE values, indicating better accuracy in predicting the target variable. The RF model also performed well with relatively low MAE, MSE, and RMSE, as shown in Figure 8.

TABLE II. COMPARISON OF EVALUATION METRICS

Model	MAE	MSE	RMSE	R^2
ARIMA	607.45	510929.19	714.79	-0.84
RF	233.801	108111.475	328.803	0.585
LSTM	125.29	55114.21	234.76	0.8012

The RF model had an R^2 value of 0.585, indicating that approximately 58.5% of the variance in the target variable could be explained by the model. The LSTM model showed a higher R^2 value of 0.8012, suggesting a better fit to the data compared to the other models. The R^2 value also signifies the goodness of fit of the model to the data. In this context, R^2 measures the proportion of variance in the target variable (crop prices) that is explained by the model. An R^2 value of 1.0 would indicate a perfect fit, which means that the model can explain 100% of the variance in the data. An R^2 value of 0.8012 is considered quite high, which indicates that the LSTM model has a strong ability to capture the underlying patterns and dependencies in the crop price data.

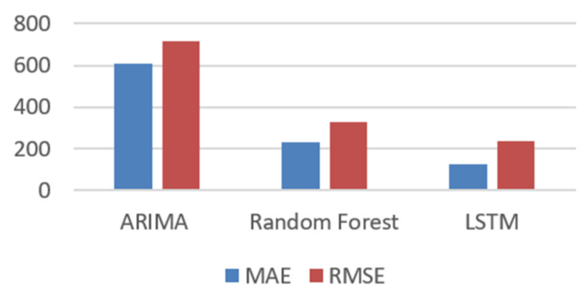

Fig. 9. Comparative analysis of models' performance.

This study provides a scientific application of big data frameworks to forecast agricultural commodity prices. It advances predictive methods by utilizing Apache Spark for data preprocessing and prediction, contributing to the development of more sophisticated techniques. By evaluating various forecasting models and revealing the limitations of traditional models such as ARIMA, while highlighting the strengths of deep learning techniques such as LSTM, this study improves the scientific understanding of modeling approaches for time series analysis. From a societal perspective, this study empowers farmers, traders, and policymakers to make more informed decisions about agricultural commodity prices, thus promoting economic stability. Furthermore, the study acknowledges its limitations and suggests the integration of additional data sources, such as weather data and market sentiment analysis, which can further improve forecast accuracy and risk management, benefiting the wider society and economy.

IV. CONCLUSION

This paper presented a novel use of a big data framework to predict the prices of agricultural commodities. The Apache Spark framework was used for data preprocessing and the prediction of crop prices in different cities from 2007 to 2022. The price of the spinach crop in Vehari City was analyzed using ARIMA, Random Forest, and LSTM. The evaluation of different models for the prediction task yielded valuable insights. The ARIMA model, although widely used for time series analysis, failed to capture the nonlinear patterns present in the data, leading to the exploration of alternative approaches. The Random Forest model demonstrated improvement by successfully capturing nonlinear patterns, but still failed to achieve accurate results. However, the LSTM model emerged as the standout performer, perfectly replicating the test data and showcasing its ability to capture complex temporal dependencies. The success of the LSTM model highlights the efficiency of deep learning techniques, particularly for time series forecasting tasks. Further refinement and exploration of deep learning models hold promise for enhancing predictive accuracy in similar domains. These results emphasize the importance of iterative analysis, model evaluation, and adaptation to select the most suitable approach to achieve accurate and reliable predictions. By harnessing the power of big data, farmers, traders, and policymakers can make more informed decisions, leading to better economic outcomes and increased stability in commodity prices.

Although this study contributes valuable insights into crop price forecasting, it is essential to acknowledge its limitations. The accuracy of the forecasting models can be influenced by external factors, such as sudden policy changes or unforeseen events, which were not explicitly considered in this analysis. The availability of historical data can vary between different regions and commodities, potentially affecting the generalizability of the findings. Additionally, the model performance may depend on the specific choice of hyperparameters and data preprocessing techniques, which could be further fine-tuned for optimization. Future research can explore the integration of additional data sources, such as weather data and market sentiment analysis, to improve the accuracy of price forecasts and provide a comprehensive understanding of the factors influencing agricultural commodity prices.

REFERENCES

[1] Y. H. Gu, D. Jin, H. Yin, R. Zheng, X. Piao, and S. J. Yoo, "Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM," *Agriculture*, vol. 12, no. 2, Feb. 2022, Art. no. 256, https://doi.org/10.3390/agriculture12020256.

[2] "The impact of higher agricultural commodity prices on emerging and low-income countries," *CaixaBank Research*, Dec. 22, 2022. https://www.caixabankresearch.com/en/economics-markets/financial-markets/impact-higher-agricultural-commodity-prices-emerging-and-low.

[3] "Agriculture Statistics", *Pakistan Bureau of Statistics*, https://www.pbs.gov.pk/content/agriculture-statistics.

[4] "Pakistan Arable Land 1961-2023." https://www.macrotrends.net/countries/PAK/pakistan/arable-land.

[5] S. Škrbić *et al.*, "Analysis of Plant-Production-Obtained Biomass in Function of Sustainable Energy," *Sustainability*, vol. 12, no. 13, Jan. 2020, Art. no. 5486, https://doi.org/10.3390/su12135486.

[6] U. Ali *et al.*, "Climate change impacts on agriculture sector: A case study of Pakistan," *Ciência Rural*, vol. 51, Apr. 2021, Art. no. e20200110, https://doi.org/10.1590/0103-8478cr20200110.

[7] N. C. Eli-Chukwu, "Applications of Artificial Intelligence in Agriculture: A Review," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4377–4383, Aug. 2019, https://doi.org/10.48084/etasr.2756.

[8] A. Rasheed, M. S. Younis, F. Ahmad, J. Qadir, and M. Kashif, "District Wise Price Forecasting of Wheat in Pakistan using Deep Learning." arXiv, Mar. 05, 2021, https://doi.org/10.48550/arXiv.2103.04781.

[9] A. Qaiser, M. U. Farooq, S. M. N. Mustafa, and N. Abrar, "Comparative Analysis of ETL Tools in Big Data Analytics," *Pakistan Journal of Engineering and Technology*, vol. 6, no. 1, pp. 7–12, Jan. 2023, https://doi.org/10.51846/vol6iss1pp7-12.

[10] S. M. Nabeel Mustafa, M. Umer Farooque, M. Tahir, S. M. Khan, and R. Qamar, "Frameworks, Applications and Challenges in Streaming Big Data Analytics: A Review," in *2022 3rd International Conference on Innovations in Computer Science & Software Engineering (ICONICS)*,

Karachi, Pakistan, Sep. 2022, pp. 1–6, https://doi.org/10.1109/ICONICS56716.2022.10100410.

[11] "Apache Spark™ - Unified Engine for large-scale data analytics." https://spark.apache.org/.

[12] A. Mishra, "Machine learning classification models for detection of the fracture location in dissimilar friction stir welded joint," *Applied Engineering Letters*, 2020.

[13] A. Rehman, Z. Deyuan, I. Hussain, M. S. Iqbal, Y. Yang, and L. Jingdong, "Prediction of Major Agricultural Fruits Production in Pakistan by Using an Econometric Analysis and Machine Learning Technique," *International Journal of Fruit Science*, vol. 18, no. 4, pp. 445–461, Oct. 2018, https://doi.org/10.1080/15538362.2018.1485536.

[14] N. Khan *et al.*, "Prediction of Oil Palm Yield Using Machine Learning in the Perspective of Fluctuating Weather and Soil Moisture Conditions: Evaluation of a Generic Workflow," *Plants*, vol. 11, no. 13, Jan. 2022, Art. no. 1697, https://doi.org/10.3390/plants11131697.

[15] B. Wang *et al.*, "Research on Hybrid Model of Garlic Short-term Price Forecasting based on Big Data," *Computers, Materials & Continua*, vol. 57, no. 2, pp. 283–296, 2018, https://doi.org/10.32604/cmc.2018.03791.

[16] S. Akshay Prassanna *et al.*, "Crop value forecasting using decision tree regressor and models," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 2, 2020.

[17] A. Vohra, N. Pandey, and S. K. Khatri, "Decision Making Support System for Prediction of Prices in Agricultural Commodity," in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, Oct. 2019, pp. 345–348, https://doi.org/10.1109/AICAI.2019.8701273.

[18] S. Bayona-Oré, R. Cerna, and E. Tirado Hinojoza, "Machine Learning for Price Prediction for Agricultural Products," *WSEAS Transactions on Business and Economics*, vol. 18, pp. 969–977, Jun. 2021, https://doi.org/10.37394/23207.2021.18.92.

[19] S. A. Haider *et al.*, "LSTM Neural Network Based Forecasting Model for Wheat Production in Pakistan," *Agronomy*, vol. 9, no. 2, Feb. 2019, Art. no. 72, https://doi.org/10.3390/agronomy9020072.

[20] Y. Su and X. Wang, "Innovation of agricultural economic management in the process of constructing smart agriculture by big data," *Sustainable Computing: Informatics and Systems*, vol. 31, Sep. 2021, Art. no. 100579, https://doi.org/10.1016/j.suscom.2021.100579.

[21] "AMIS Agriculture Marketing Wing Punjab." http://www.amis.pk/.

[22] S. K. Filipova-Petrakieva and V. Dochev, "Short-Term Forecasting of Hourly Electricity Power Demand: Reggresion and Cluster Methods for Short-Term Prognosis," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8374–8381, Apr. 2022, https://doi.org/10.48084/etasr.4787.

[23] S. Joseph, N. Mduma, and D. Nyambo, "A Deep Learning Model for Predicting Stock Prices in Tanzania," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10517–10522, Apr. 2023, https://doi.org/10.48084/etasr.5710.