Black Sea Journal of Agriculture
doi: 10.47115/bsagriculture.1304625

Open Access Journal
e-ISSN: 2618 – 6578

Research Article
Volume 6 - Issue 4: 422-426 / July 2023

COMPARATIVE ANALYSIS OF CNN, LSTM AND RANDOM FOREST FOR MULTIVARIATE AGRICULTURAL PRICE FORECASTING

Cevher ÖZDEN[1]*

[1]*Mugla Sitki Kocman University, Faculty of Bodrum Fine Arts, Department of Digital Game Design, 48420, Mugla, Türkiye*

Abstract: Time series forecasting is an important research topic among agriculture economics. Especially, multivariate, multi-step and multiple output prediction tasks pose a challenge in research as their nature requires the investigation of intra- and inter-series correlation. The common statistical methods like ARIMA and SARIMA fall short in this kind of tasks. Deep learning architectures like Convolutional Neural Networks and Long Short-Term Memory networks are quite good at modelling the structures of complex data relations. In this study, a new dataset is composed through manual collection of data from the Ministry of Commerce of Turkish Republic. The dataset contains daily trade volumes and prices of potato, onion and garlic, which are most commonly consumed products in Turkish cuisine. The data pertains to the period between January 1, 2018 and November 26, 2022 (1791 days). A simple CNN and LSTM architectures as well Random Forest machine learning method are used to predict the next 10-day prices of the products. Accordingly, three models provided acceptable results in the prediction tasks, while CNN yielded by far the best result (MAE: 0.047, RMSE: 0.070).

Keywords: Time series, CNN, LSTM, Price prediction, Agriculture

***Corresponding author:** Mugla Sitki Kocman University, Faculty of Bodrum Fine Arts, Department of Digital Game Design, 48420, Bodrum, Mugla, Turkiye
E mail: efeozden@gmail.com (C. ÖZDEN)
Cevher ÖZDEN https://orcid.org/0000-0002-8445-4629

Received: May 28, 2023
Accepted: June 23, 2023
Published: July 01, 2023

Cite as: Özden C. 2023. Comparative analysis of CNN, LSTM and random forest for multivariate agricultural price forecasting. BSJ Agri, 6(4): 422-426.

1. Introduction

Natural conditions have a significant impact on agricultural production because of its inherent qualities. Farmers must deal with risks and uncertainties in the production process as a natural result of this circumstance (Skendžić et al., 2021). Production is specifically significantly impacted by unfavorable meteorological events, illnesses, pests, and price uncertainty (Pham et al., 2021). Farmers, producers, and other agricultural industry players' decision-making processes are impacted by price uncertainties in agricultural output. These uncertainties are linked to the erratic and unpredictable nature of agricultural commodity price movements (Molitor et al., 2017). Weather conditions, political and economic variables, diseases and pests, global marketplaces, and the market's supply-demand dynamics are all elements that affect price uncertainty in agriculture (Mili and Bouhaddane, 2021).

The predictability of the future price change of these products is very important for the producers and consumers, as well as for the intermediaries, exporting and importing countries, because the prices of agricultural products can affect the general economic situation not only in the national but also in the international market (Duarte et al., 2021). Prices fluctuate abruptly as a result of shifts in supply and demand for agricultural goods, which has a direct impact on people's daily life. Accurately predicting the future price of a commodity aids farmers and consumers in being aware of the risks posed by price variations, taking into account how changes in the pricing of agricultural products also impact supply and demand. (Zou et al., 2022).

Time series analysis is a statistical technique that is used to look at how data changes over time and forecast future values. Data that is measured and organized in time series at regular intervals (Nielsen, 2020). Future value forecasting, trend identification, seasonal variation understanding, cycle identification, and other patterns of change across time are all accomplished through time series analysis (Jebb et al., 2015). These days, a variety of fields employ this technique to generate future forecasts. This technique, for instance, is used to forecast market prices (Dingli and Fournier, 2017; Yadav et al., 2020), to estimate air quality (Freeman et al., 2018; Espinoza et al., 2021), and to forecast the number of deaths during the Covid-19 outbreak (Kırbaş et al., 2020; Shastri et al. 2020).

Making critical judgments is made easier when the course of events or metrics can be properly predicted. by this vantage point, all parties involved in this industry can profit by being able to forecast the prices of agricultural products. Using historical datasets to make predictions is challenging because it necessitates modeling both intra-series temporal models and inter-series correlations jointly. Analyzing multivariate time series with several steps and outputs can be challenging. The ability to capture inter-series correlations and provide various outputs is one of the limitations of conventional statistical analysis techniques like ARIMA and SARIMAX. The development of deep learning algorithms has allowed for the use of fresh viewpoints in the solution of this issue. The most extensively studied neural network models in this regard include LSTM and CNN. The architectural layouts of the LSTM model allow it to accurately forecast sequence pattern information. The final prediction model will benefit more from the ability of the CNN model to filter out noise in the input data and extract more valuable features (Casado-Vara et al., 2021). In contrast to standard CNNs, which are well suited to handle spatial autocorrelation data but are typically not adapted to correctly manage complex and long temporal dependencies, LSTM networks are designed to handle temporal correlations but only use the features that are present in the training set (Bengio et al., 2013).

In this study, future price predictions of 3 basic foodstuffs such as potatoes, onions and garlic were compared using deep learning methods. Potatoes, onions and garlic are important products for staple food consumption in Türkiye. These products play an important role in the nutrition of the people as basic foodstuffs. They contain high amounts of vitamins, minerals and other nutrients and provide the components necessary for a balanced diet. Therefore, the production of potatoes, onions and garlic is of great importance in terms of maintaining domestic food production and ensuring food security. They are mostly used together in preparing daily meals in traditional Turkish cuisine and constitute the main ingredients of daily nutrition of large part of Turkish people. Therefore, any price change in these products has direct effects on family budget in Türkiye. Also, their trade volumes and prices are interconnected and effective on each other. In this study, we have curated daily trade volumes and prices of potato, onion and garlic from the official market website of Turkish Ministry of Commerce (www.hal.gov.tr). The resulting dataset requires multivariate input, multi-step and multiple output time series analysis. For this purpose, LSTM, CNN and RandomForest methods are applied to forecast future prices of the products. The main contribution of the study is that a new dataset is made publicly available for scientists who would like to experiment on agricultural product price analysis. Another important contribution is that policy makers and traders can use the results while taking their decisions in trade.

Agricultural commodity futures price forecasting is a crucial topic in the agricultural sector since it helps to reduce market uncertainty and risks by not only giving decision-makers with accurate price information for agricultural commodities in advance (Bayona-Oré,2021). In this context, it is noteworthy that various studies have been carried out by using time series and deep learning algorithms in price predictions of agricultural products.

Madaan et al. (2019) forecasted the price of potato and onion goods in India using the ARIMA, SARIMA, and LSTM models. The research's findings showed that the LSTM model performed better than the ARIMA and SARIMA models. A model was developed to predict mandi prices over a 30-day period, and it provided an average RMSE value of 754.6 (or 25.15 for a single day) over those periods. With a mean normalized deviation of 0.041, the performance was acceptable. This approach might assist farmers in determining whether to offer price forecast data, hold onto their produce for a few weeks, or sell it right away.

In order to forecast the price of squash in Sri Lanka, Navaratnalingam et al. (2020) used a deep learning system called the LSTM and ARIMA model. The research's findings showed that the multivariate CNN LSTM model outperformed other models, offering an average RMSE of 19.46 Sri Lankan rupees per kilogram with an average RMSPE of 14.9%. The analysis also revealed that there is a connection between price variation and typical days of the week.

Dharavath and Khosla (2019) used ARIMA and SARIMA models to predict the price of fruits such as Mango and Pineapple. Although the researchers tested various models on the data they obtained, they could not obtain 100% accurate results. As the most important factor in this, they emphasized the necessity of providing regular data for time series analysis.

2. Materials and Methods

In this study, daily trade volumes and prices of onion, potato and garlic are collected from the official site of Marketplace Registration System, which is affiliated to the Ministry of Commerce of Türkiye. The resulting data pertains to the period between January 1, 2018 and November 26, 2022 (1.791 days). The curated dataset, analysis results and codes are made publicly available in a Github repository (https://github.com/cevher/Potato-Onion-Garlic-Time-Series).

A quite basic CNN architecture is applied to forecast multi target price outputs in the study. The model contains Conv1D (kernel size=2), MaxPooling1D (pool_size=2), Conv1D (kernel_size=3), MaxPooling1D, Flatten, Dense (activation=ReLU), Dense (output layer) with adam optimizer and 'mse' loss function. The network is trained for 200 epochs. A quite simple architecture is chosen for LSTM network, which contains LSTM and Dense layers with adam optimizer and 'mse' loss function. This network is trained for 200 epochs, as well.

3. Results

The data consists of daily trade volumes of prices of potato, onion and garlic which are the main ingredients of Turkish cuisine. The data pertains to the period between January 2st 2018 and November 6th 2022 and contains no missing value. Prior to analysis, min-max scaler is used to standardize the values in order to overcome any bias among data. Outline data are cleaned and first difference is taken to establish non-stationarity. The pre- and post-states of the dataset are given in Figure 1.

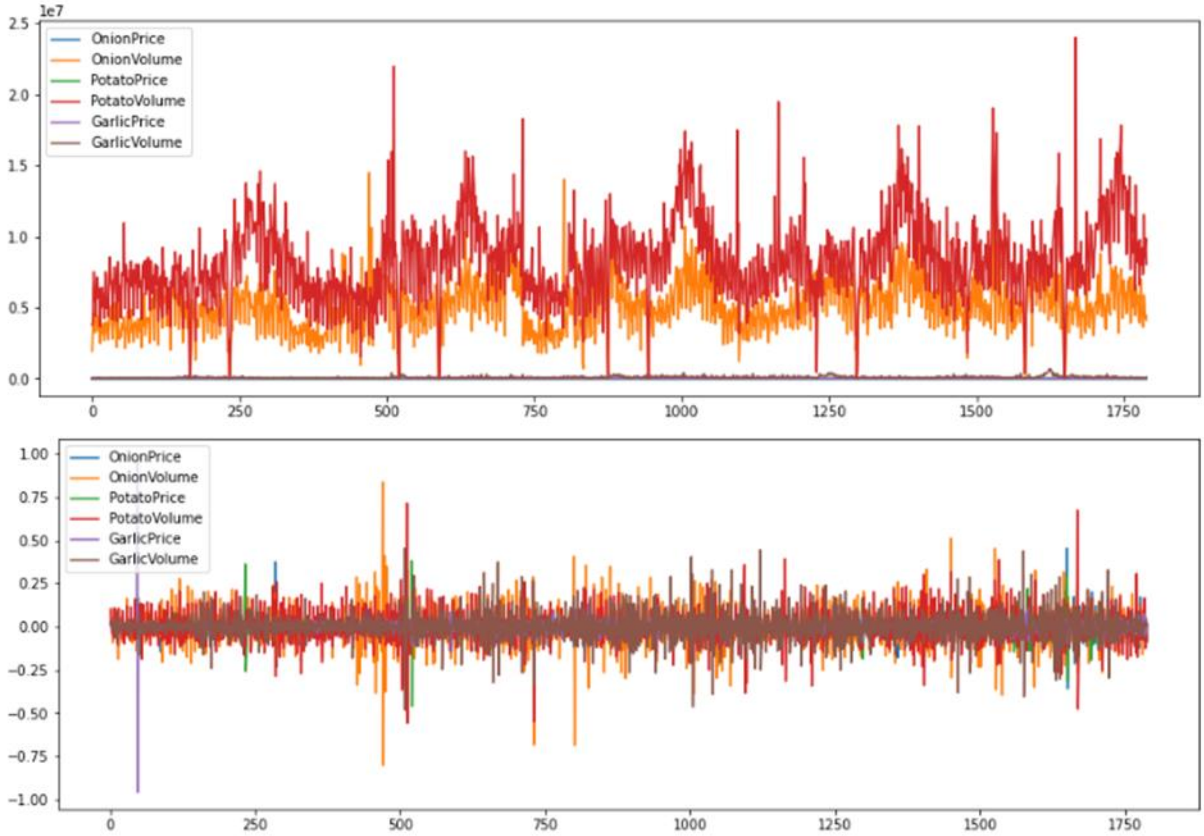

Figure 1. Pre- and post-state of data following the preprocessing.

Data is split into train (90%) and test sets (10%). Subsequently, 6 columns of 28 days containing price and trade volume of three products are considered for input and 3 price columns of three products are forecasted for the next 10 days. Models are run in a computer with the following specifications (Intel(R) Core(TM) i7-7500U CPU, 2.70GHz). Model results are evaluated through Mean Absolute Error (MAE) and Root Mean Square Error (RMSA), which are common methods used in time series analysis. The performance is considered better with lower metric scores. The metric scores of the evaluated models are given in Table 1. Accordingly, all three models provided quite good results in predicting prices for the next 10 days, while CNN clearly obtained by far the best results in both metrics.

Table 1. Forecasting results for 5 days and 10 days

	10 days forecast	
	MAE	RMSE
Random Forest	3.953	6.0797
CNN	0.047	0.070
LSTM	3.837	5.929

4. Discussion

The results indicate that CNN models detected the inter- and intra-series correlations among data points. Similar results have been reported by the previous researches on wheat yield prediction, oil production prediction and nonlinear time history prediction of seismic responses. The results of the current study support the previous findings. Wang et al. (2018) compared the performance of CNN models with traditional time series models for agricultural price forecasting. The results demonstrated that CNNs can effectively capture spatial dependencies in multivariate time series data and outperform traditional models like ARIMA and SVM. Wibawa et al. (2022) discussed the successful application of CNNs in various time series forecasting tasks, including agricultural price prediction. The research highlighted the ability of CNNs to automatically learn relevant features and patterns from multivariate time series data, leading to improved forecasting accuracy. Torres et al. (2020) provided an overview of deep learning techniques for time series forecasting, with LSTM being one of the prominent methods. The study showcased the effectiveness of LSTM models in capturing long-term dependencies and demonstrated their superior performance compared to traditional time series models in various domains, including agriculture. Klompenburg et al. (2020) discussed the application of various machine learning algorithms, including Random Forests, for crop yield prediction. Although not solely focused on price forecasting, the review indicated that Random Forest models could be adapted to predict agricultural prices based on relevant features.

5. Conclusion

This study makes a comparative investigation into the applicability of three different AI methods (CNN, LSTM and Random Forest) on multivariate, multi-step and multiple output prediction task. The dataset contains daily price and trade volumes of three most commonly used agricultural products (onion, potato, garlic) in Türkiye. These products are the main ingredients of Turkish dietary and their prices have direct effects on food expenditures of a large segment of population. The models have been tested on 10 days' price forecast tasks for three products at the same run, and their results are compared with each other. Accordingly, all three models yielded good results in prediction tasks; however, CNN model clearly outperformed other models. CNN, RandomForest and LSTM models have been commonly used for prediction tasks in literature. Many of the studies employed deep architectures of the models consisting of large number of layers. In the current study, a quite basic form of models are considered. Despite their low number of layers, all models are concluded to be good alternatives for prediction tasks in agriculture. Further studies can be carried out using combination of CNN, LSTM, GRU networks with more layers.

Author Contributions

The percentage of the author contributions is present below. The author reviewed and approved final version of the manuscript.

	C.Ö.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Bayona-Oré S, Cerna R, Tirado Hinojoza E. 2021. Machine learning for price prediction for agricultural products. Wseas Transact Busin Econ, URL: https://doi.org/10.37394/23207.2021.18.92 (accessed date: March 11, 2022).

Bengio Y, Courville A, Vincent P. 2013. Representation learning: A review and new perspectives. IEEE Transact Pattern Anal Machine Intell, 35(8): 1798-1828. https://doi.org/10.1109/TPAMI.2013.50.

Casado-Vara R, Martin del Rey A, Pérez-Palau D, de-la-Fuente-Valentín L, Corchado J. M. 2021. Web traffic time series forecasting using LSTM neural networks with distributed asynchronous training. Math, 9(4): 421. https://doi.org/10.3390/math9040421.

Dharavath R, Khosla E. 2019. Seasonal ARIMA to forecast fruits and vegetable agricultural prices. In Proceedings of the IEEE International Symposium on Smart Electronic Systems, 16-18 December, Rourkela, India, pp: 47-52.

Dingli A, Fournier KS. 2017. Financial time series forecasting-a deep learning approach. Inter J Mach Learn Comput, 7(5): 118-122.

Duarte AMA, Gaglianone WP, de Carvalho Guillén OT, Issler JV. 2021. Commodity prices and global economic activity: a derived-demand approach. Energy Econ, 96: 105120. https://doi.org/10.1016/j.eneco.2021.105120.

Espinosa R, Palma J, Jiménez F, Kamińska J, Sciavicco G, Lucena-Sánchez E. 2021. A time series forecasting based multi-criteria methodology for air quality prediction. Appl Soft Comput, 113: 107850. https://doi.org/10.1016/j.asoc.2021.107850.

Freeman BS, Taylor G, Gharabaghi B, Thé J. 2018. Forecasting air quality time series using deep learning. J Air Waste Manage Assoc, 68(8): 866-886. https://doi.org/10.1080/10962247.2018.1459956.

Jebb AT, Tay L, Wang W, Huang Q. 2015. Time series analysis for psychological research: examining and forecasting change. Front Psychol, 6: 727.

Klompenburg T, Kassahun A, Catal C. 2020. Crop yield prediction using machine learning: A systematic literature review. Comput Electron Agri, 177. 105709. 10.1016/j.compag.2020.105709.

Kırbaş İ, Sözen A, Tuncer A. D, Kazancıoğlu F. Ş. 2020. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA NARNN and LSTM approaches. Chaos Solitons Fractals, 138: 110015. https://doi.org/10.1016/j.chaos.2020.110015.

Madaan L, Sharma A, Khandelwal P, Goel S, Singla P, Seth A. 2019. Price forecasting anomaly detection for agricultural commodities in India. In Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies, July 3 – 5, Accra, Ghana, pp: 52-64.

Mili S, Bouhaddane M. 2021. Forecasting global developments and challenges in olive oil supply and demand: A Delphi survey from Spain. Agri, 11(3): 191. https://doi.org/10.3390/agriculture11030191.

Molitor K, Braun B, Pritchard B. 2017. The effects of food price changes on smallholder production and consumption decision-making: evidence from Bangladesh. Geograp Res, 55(2): 206-216. https://doi.org/10.1111/1745-5871.12225.

Navaratnalingam S, Kodagoda N, Suriyawansa K. 2020. Exploiting multivariate LSTM models with multistep price forecasting for agricultural produce in Sri Lankan context. In Proceedings of 2nd International Conference on Advancements in Computing (ICAC), 10-11 December 2020, Colombo, Sri Lanka, pp: 328-332.

Nielsen A. 2020. Practical Time Series Analysis: Prediction with Statistics and Machine Learning; O'Reilly: Sebastopol CA USA.

Pham TT, Dang HL, Pham NTA, Dang HD. 2021. Adoption of contract farming for managing agricultural risks: A case study in rice production in the Mekong Delta Vietnam. J Agribus Develop Emerg Econ, 0839. https://doi.org/10.1108/JADEE-05-2021-0107.

Shastri S, Singh K, Kumar S, Kour P, Mansotra V. 2020. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. Chaos Solitons Fractals, 140: 110227. https://doi.org/10.1016/j.chaos.2020.110227.

Skendžić S, Zovko M, Živković IP, Lešić V, Lemić D. 2021. The impact of climate change on agricultural insect pests. Insects 12(5): 440. https://doi.org/10.3390/insects12050440.

Torres J, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. 2020. Deep learning for time series forecasting: A survey. Big Data, 9: 3-21. 10.1089/big.2020.0159.

Wang K, Li K, Zhou L, Hu Y, Zhongyao C, Liu J, Chen C. 2019. Multiple convolutional neural networks for multivariate time series prediction. Neurocomputing. 360: 107-119. 10.1016/j.neucom.2019.05.023.

Wibawa A, Putra UA, Elmunsyah H, Pujianto U, Dwiyanto F, Hernandez L. 2022. Time-series analysis with smoothed Convolutional Neural Network. J Big Data, 9: 44. 10.1186/s40537-022-00599-y.

Yadav A, Jha CK, Sharan A. 2020. Optimizing LSTM for time series prediction in Indian stock market. Procedia Comput Sci, 167: 2091-2100. https://doi.org/10.1016/j.procs.2020.03.257.

Zou G, Guo Z Sun S, Jin Q. 2023. A CNN-BiGRU-AM neural network for AI applications in shale oil production prediction. Appl Ener, 344: 121249.