Predicting Food Supply with LSTM: A Data-Driven Approach Using WASDE Data

I Made Dwi Jendra Sulastra[1], I Putu Oka Wisnawa[2], I Komang Wiratama[3], and I Putu Astya Prayudha[4]

[1] Accounting Department, Politeknik Negeri Bali, Bali, Indonesia
[2,3,4] Information Technology Department, Politeknik Negeri Bali, Bali, Indonesia
dwijendrasulastra@pnb.ac.id

Abstract. Ensuring a stable and sustainable food supply is a critical global challenge. Accurate forecasting of food supply is essential for effective policymaking, resource allocation, and risk management. This study proposes a novel approach to food supply forecasting using a Long Short-Term Memory (LSTM) neural network model, leveraging the World Agricultural Supply and Demand Estimates (WASDE) dataset. The study aims to develop a robust LSTM model capable of accurately predicting food supply and evaluating its performance using appropriate metrics. The LSTM model was trained on historical WASDE data, which includes information on production, consumption, and ending stocks for various commodities. The model's performance was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The results indicate that the LSTM model achieved varying levels of success in predicting food supply for different commodities. For some commodities, such as rice and coarse grain, the model demonstrated strong accuracy with low RMSE and MAE values. However, for other commodities, like corn and wheat, the model struggled to accurately predict supply, especially during periods of high volatility. In conclusion, this study demonstrates the potential of LSTM models in food supply forecasting, utilizing the WASDE dataset. While the model achieved promising results for certain commodities, further research is needed to improve its accuracy for more challenging commodities. Future studies could explore incorporating additional factors, such as climate change and geopolitical events, to enhance the model's predictive capabilities.

Keywords: Food Supply Forecasting, LSTM, Time Series Analysis, WASDE Data

1 Introduction

Ensuring a stable and sustainable food supply is a critical global challenge. Accurate forecasting of food supply is essential for effective policymaking, resource allocation, and risk management. This study proposes a novel approach to food supply forecasting using a Long Short-Term Memory (LSTM) neural network model, leveraging the World Agricultural Supply and Demand Estimates (WASDE) dataset.

A. A. N. G. Sapteka et al. (eds.), *Proceedings of the International Conference on Sustainable Green Tourism Applied Science - Engineering Applied Science 2024 (ICoSTAS-EAS 2024)*, Advances in Engineering Research 249,
https://doi.org/10.2991/978-94-6463-587-4_44

LSTM networks, renowned for their ability to capture long-term dependencies in time series data, have demonstrated remarkable performance in various forecasting tasks (Mehtab & Sen, 2020; Althelaya et al., 2018; Falatouri et al., 2022; Zuo et al., 2022; Barzegar et al., 2020; Fischer & Krauss, 2018). By utilizing the WASDE dataset, which provides comprehensive information on global agricultural supply and demand, this study aims to develop a robust and accurate food supply forecasting model.

The research aims to achieve two primary objectives. First, construct a robust LSTM-based model capable of accurately predicting food supply. The model will be trained on historical WASDE data, which provides comprehensive information on global agricultural supply and demand (Roznik et al., 2023; Ding & Katchova, 2023; Adjemian et al., 2018). By leveraging the LSTM architecture, the model will be able to capture complex patterns and dependencies in the data, leading to more accurate predictions. Then, the second is to assess the performance of the developed LSTM model using appropriate evaluation metrics. This evaluation will involve comparing the model's predictions to actual food supply data to determine its accuracy and robustness.

This study contributes to the growing body of research on food security and sustainable agriculture. By demonstrating the potential of LSTM models and the WASDE dataset in food supply forecasting, we aim to provide valuable insights for policy-makers, researchers, and industry stakeholders.

2 Methodology

This research employed a Quantitative Predictive Design utilizing a Long Short-Term Memory (LSTM) algorithm to forecast global food supply based on WASDE data. LSTM was selected due to its demonstrated robustness and accuracy in time series forecasting in previous studies (Mehtab & Sen, 2020; Althelaya et al., 2018; Falatouri et al., 2022; Zuo et al., 2022; Barzegar et al., 2020; Fischer & Krauss, 2018). Comparisons to other methods have consistently shown LSTM's superior performance (Navares & Aznarte, 2020; Wang et al., 2020). While LSTM can be combined with other methods to enhance performance, this study focuses solely on LSTM to provide a baseline assessment of its capabilities in time series forecasting.

Secondary data was obtained from the official WASDE website. The dataset comprised approximately 827,869 data points, spanning from April 2010 to June 2024. Data preprocessing involved several filtering steps. Initially, data was filtered based on region, selecting the global region to reduce the dataset to 33,774 rows. Subsequently, data was filtered based on commodity attributes. Given the diversity of commodities in the WASDE dataset, attributes such as production, exports, domestic use, traded quantity, and ending stocks were considered. To ensure data consistency, attributes commonly found across most commodities production, exports, and ending stocks were selected, resulting in approximately 164 data points per commodity. This refined dataset was then divided into training and testing sets in an 80:20 ratio, adhering to Pareto's principle.

Table 1. Amount of data used as dataset

Commodity	Total dataset	Data training	Data testing
Rice	164	131	33
Corn	164	131	33
Coarse Grain	164	131	33
Cotton	164	131	33
Soybean Meal	164	131	33
Soybean Oil	164	131	33
Wheat	164	131	33

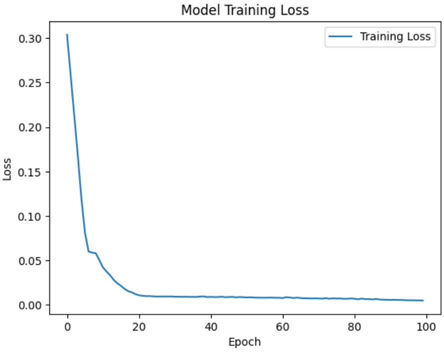
3 Result and Discussion

3.1 Result

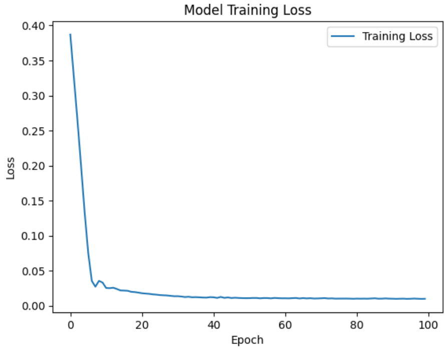
In this research, the LSTM model was trained for 100 epochs for each commodity dataset. Figure 1 illustrates the performance of the proposed and tested LSTM model. Loss is a measure of the failure of a method or algorithm in completing a given task. In this case, the task is to predict the global food supply based on the provided WASDE data. Initially, at the beginning of the epoch, the loss value was relatively high, as loss values typically range from 0 to 1. However, this loss value decreased as the number of epochs increased. For almost all commodities, the loss value began to decrease and stabilize around the 20th epoch.

The results of the experiment implementing LSTM on WASDE data for various commodities show varying levels of performance. For rice, the model achieved a low loss of 0.00451608, indicating a good model fit. The RMSE of 2.23828232 and MAE of 1.59488482 suggest that the model's predictions were relatively accurate.

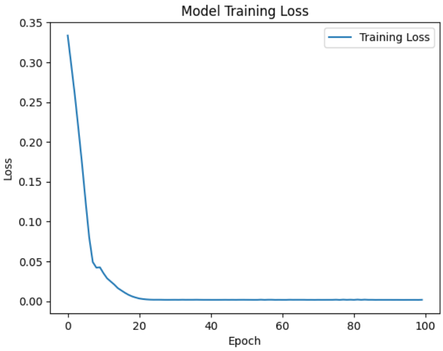
For corn, the model's performance was less impressive, with a higher loss of 0.00791964 and significantly larger RMSE and MAE values. This indicates that the model struggled to accurately predict corn supply. For coarse grain, the model demonstrated strong performance with a low loss of 0.00140588 and reasonable RMSE and MAE values. Cotton also showed good results, with a low RMSE and MAE, suggesting accurate predictions. Soybean meal and soybean oil had mixed results. Soybean meal had a relatively high loss and RMSE, while soybean oil had a lower loss and RMSE. Wheat exhibited the highest loss and RMSE among all commodities, indicating that the model struggled to predict wheat supply accurately. The result is shown in Figure 1 and Table 2.


Model Training Loss
Training Loss
Loss
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80
100
Epoch

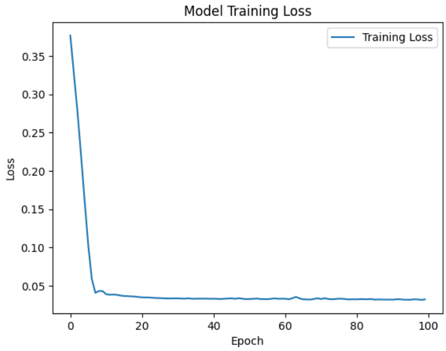
(a)


Model Training Loss
Training Loss
Loss
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80
100
Epoch

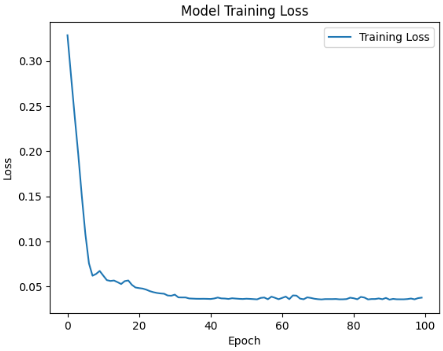
(b)


Model Training Loss
Training Loss
Loss
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80
100
Epoch

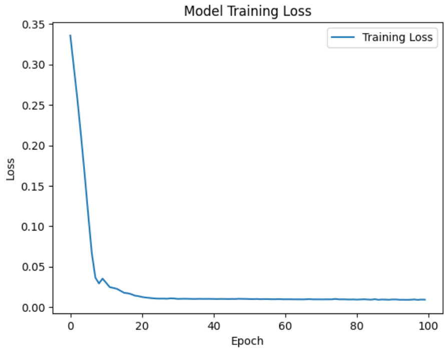
(c)


Model Training Loss
Training Loss
Loss
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
20
40
60
80
100
Epoch

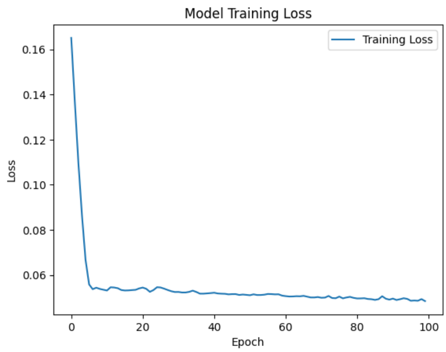
(d)


Model Training Loss
Training Loss
Loss
0.30
0.25
0.20
0.15
0.10
0.05
0
20
40
60
80
100
Epoch

(e)


Model Training Loss
Training Loss
Loss
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80
100
Epoch

(f)



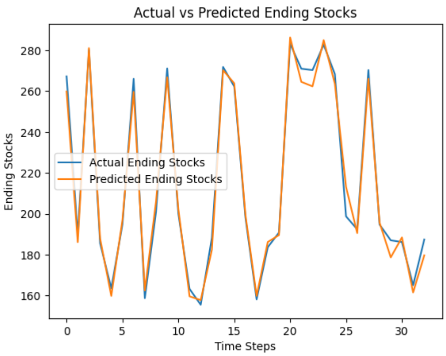
(g)

Figure 1. Rice (a), Corn (b), Coarse grain (c), Cotton (d), Soybean meal (e), Soybean oil (f), Wheat (g)

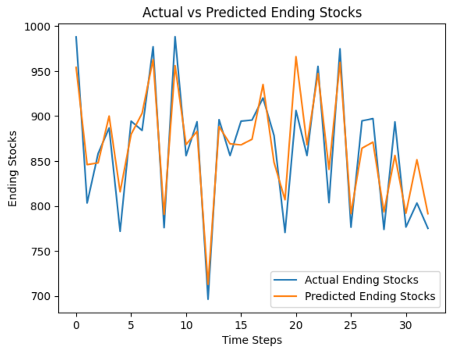
Table 2. Amount of data used as dataset

Commodity	Loss	RMSE	MAE
Rice	0.00451608	2.23828232	1.59488482
Corn	0.00791964	26.64032308	23.30351328
Coarse Grain	0.00140588	4.88344529	3.98547464
Cotton	0.02043436	1.18630779	0.92615747
Soybean Meal	0.04384283	5.10623583	4.71887035
Soybean Oil	0.00797728	0.86844531	0.69630509
Wheat	0.05427938	6.91669779	6.05186886

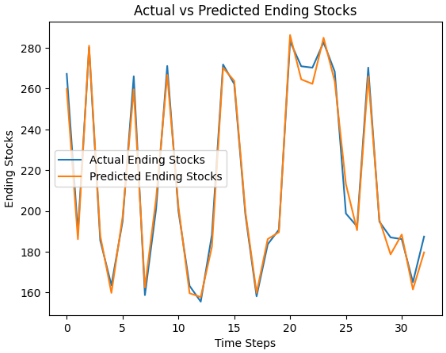
Overall, the LSTM model achieved varying levels of success in predicting commodity supply based on WASDE data. The performance of the model differed significantly across commodities, highlighting the complexity of predicting supply for different agricultural products. For rice and coarse grain, the model demonstrated strong accuracy with low loss and RMSE values. However, for corn, wheat, and soybean meal, the model struggled to accurately predict supply. Soybean oil showed mixed results, with a relatively low loss and RMSE but higher MAE. These findings highlight the complexity of predicting commodity supply and emphasize the need for further research and experimentation to improve the model's performance for all commodities. The result is shown in Figure 2.


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

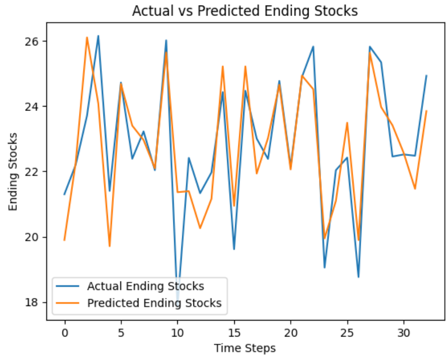
(a)


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

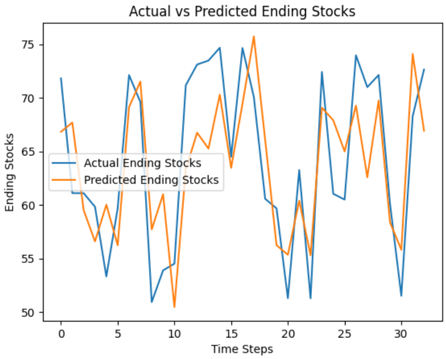
(b)


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

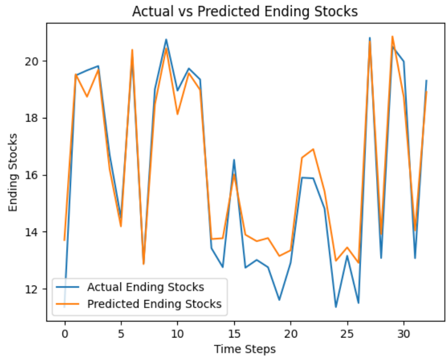
(c)


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

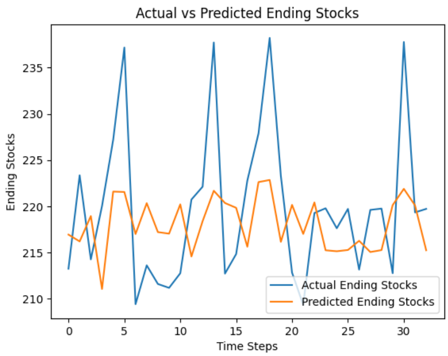
(d)


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

(e)


Actual vs Predicted Ending Stocks
Ending Stocks
Time Steps
Actual Ending Stocks
Predicted Ending Stocks

(f)



(g)

Figure 2. Rice (a), Corn (b), Coarse grain (c), Cotton (d), Soybean meal (e), Soybean oil (f), Wheat (g)

The LSTM model used in this analysis demonstrates varying levels of success in predicting ending stocks for different commodities. While it can effectively capture the general trends in the data for most commodities, there are noticeable deviations between the predicted and actual values, particularly during periods of rapid change or extreme fluctuations. This suggests that the model's ability to accurately forecast commodity ending stocks is influenced by the specific characteristics of each commodity and the complexity of its underlying dynamics. The LSTM model used appears to capture the general trends in the data for most commodities, but there are notable deviations between actual and predicted values, especially during periods of rapid change or extreme fluctuations. Each commodity exhibits unique patterns in the relationship between actual and predicted values. For instance, the model seems to perform better in predicting the ending stocks of some commodities (e.g., Rice, Coarse Grain, Cotton) compared to others (e.g., Corn, Soybean Meal).

3.2 Discussion

The LSTM model's performance in predicting ending stocks varied significantly across the different commodities analyzed. Here's a breakdown of its effectiveness for each:

Rice. The model captured the general upward or downward trend of rice ending stocks, but there were distinct discrepancies between predicted and actual values, particularly around specific time steps (10-15 and 25-30). This suggests the model may struggle with predicting short-term fluctuations in rice supply and demand.

Corn. The model exhibited significant challenges in capturing the sharp fluctuations observed in corn-ending stocks. The predicted values consistently deviated from the actual values, often overestimating or underestimating the true stock levels. This indicates that the model might not be adequately capturing the factors that cause rapid changes in corn supply and demand.

Coarse Grain. The model performed relatively well for coarse grain ending stocks. The predicted values closely tracked the actual values, suggesting the model effectively learned the underlying patterns governing coarse grain supply and demand dynamics.

Cotton. Similar to coarse grain, the model demonstrated good accuracy in predicting cotton ending stocks. The predicted values are closely aligned with the actual values, suggesting the model can effectively model the factors influencing cotton supply and demand.

Soybean Meal. The model encountered some difficulties in predicting soybean meal ending stocks, particularly during periods of rapid increase or decrease. The predicted values deviated from the actual values during these periods, indicating limitations in capturing the factors driving significant changes in soybean meal supply and demand.

Soybean Oil. The model's predictions for soybean oil ending stocks were generally accurate, with only minor deviations from the actual values. This suggests the model successfully learned the patterns influencing soybean oil supply and demand, allowing it to make relatively accurate predictions.

Wheat. The model faced significant challenges in predicting wheat-ending stocks, especially during periods of high volatility. The predicted values often deviated significantly from the actual values, indicating the model struggled to capture the complex factors driving fluctuations in wheat supply and demand.

Overall, the LSTM model exhibited varying degrees of success in predicting ending stocks for different commodities. While it could capture the general trends in most cases, it struggled with sharp fluctuations or periods of high volatility. This variability highlights the influence of specific commodity characteristics and the complexity of underlying supply and demand dynamics. Further research and experimentation are necessary to improve the model's performance, particularly for commodities exhibiting intricate patterns or influenced by external factors beyond the scope of the current model. Here are some potential areas for future research:

Data Quality. Ensuring high-quality data for training and testing is crucial. Future studies could explore incorporating additional data sources or implementing data cleaning techniques to improve data accuracy and completeness.

Model Complexity. Experimenting with different model architectures, such as increasing the number of layers or neurons in the LSTM network could potentially enhance the model's ability to capture complex patterns in the data.

External Factors. Integrating data on external factors such as weather conditions, geopolitical events and economic indicators could potentially improve the model's predictive power by accounting for their influence on commodity supply and demand.

By addressing these considerations, future research can develop more robust and accurate LSTM models for predicting commodity-ending stocks, aiding stakeholders in the agricultural sector to make informed decisions.

4 Conclusion

This research aimed to investigate the relationship between green economy strategies and food supply prediction using a Long Short-Term Memory (LSTM) model based on WASDE data. The study found that while the LSTM model can effectively capture general trends in the food supply, its performance varies significantly across different commodities. For commodities like rice and coarse grain, the model demonstrated strong accuracy in predicting ending stocks. However, for corn, wheat, and soybean meal, the model struggled to accurately forecast supply, especially during periods of high volatility. Soybean oil showed mixed results, with relatively accurate predictions for some periods and less accurate ones for others.

The findings highlight the complexity of predicting food supply and the need for further research to improve model accuracy. Future studies could explore incorporating additional external factors, such as weather conditions, geopolitical events, and economic indicators, to enhance the model's predictive power. Additionally, investigating alternative machine learning techniques or hybrid models could potentially lead to more accurate and robust predictions.

In conclusion, while the LSTM model provides valuable insights into the relationship between green economy strategies and food supply prediction, it is essential to recognize its limitations and continue exploring innovative approaches to address the challenges associated with food security in a rapidly changing world.

References

Adjemian, M. K., Johansson, R., McKenzie, A., & Thomsen, M. (2018). Was the missing 2013 WASDE missed? *Applied Economic Perspectives and Policy*, *40*(4). https://doi.org/10.1093/aepp/ppx049.

Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018). Evaluation of bidirectional LSTM for short and long-term stock market prediction. *2018 9th International Conference on Information and Communication Systems, ICICS 2018*, *2018-Janua*. https://doi.org/10.1109/IACS.2018.8355458.

Barzegar, R., Aalami, M. T., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, *34*(2). https://doi.org/10.1007/s00477-020-01776-2.

Ding, K., & Katchova, A. L. (2023). Testing the optimality of USDA's WASDE forecasts under unknown loss. *Agribusiness*. https://doi.org/10.1002/agr.21850.

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive analytics for demand forecasting - A comparison of SARIMA and LSTM in retail SCM. *Procedia Computer Science*, *200*. https://doi.org/10.1016/j.procs.2022.01.298.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2). https://doi.org/10.1016/j.ejor.2017.11.054.

Mehtab, S., & Sen, J. (2020). Stock price prediction using CNN and LSTM-based deep learning models. *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*. https://doi.org/10.1109/DASA51403.2020.9317207.

Navares, R., & Aznarte, J. L. (2020). Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*, *55*. https://doi.org/10.1016/j.ecoinf.2019.101019.

Roznik, M., Mishra, A. K., & Boyd, M. S. (2023). Using a machine learning approach and big data to augment WASDE forecasts: Empirical evidence from US corn yield. *Journal of Forecasting*, *42*(6). https://doi.org/10.1002/for.2956.

Wang, P., Zheng, X., Ai, G., Liu, D., & Zhu, B. (2020). Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru, and Iran. *Chaos, Solitons and Fractals*, *140*. https://doi.org/10.1016/j.chaos.2020.110214.

Zuo, M., Wang, Y., Yan, W., & Zhang, Q. (2022). Food enterprises- profit growth rate prediction based on LSTM from the perspective of the supply chain. *Advances in Transdisciplinary Engineering*, *30*. https://doi.org/10.3233/ATDE221102.