Proceedings of the 5th Indian International Conference on Industrial Engineering and Operations Management, Vellore, India, November 6–8, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250140
Published: November 6, 2025

A Deep Learning Ensemble Approach with Agronomic Weather Integration for District-Level Crop Price Prediction in Tamil Nadu

Manjula R
Professor, SCSE, VIT University
Vellore, Tamil Nadu, India
rmanjula@vit.ac.in

Siddhanth M K, Mohammed A K and Tanishq A
Computer Science Engineer, SCOPE, VIT University
Vellore, Tamil Nadu, India
siddhanth.monnappa2022@vitstudent.ac.in, mohammedaiman.khan2022@vitstudent.ac.in, tanishq.anand2022@vitstudent.ac.in

Abstract

This study focuses on sophisticated deep learning ensemble model to overcome the challenges faced by volatile agricultural prices in Tamil Nadu, India for 23 crops. The model provides a method to accurately predict future crop prices for market stability and economic planning. Integration of historical price from the Agmarknet portal (2015-2025) with district specific lookback weather data using the Open-Meteo API, to match the unique agronomic timeline of each crop. By establishing a clear baseline prediction using the traditional SARIMAX model and improving the results with four deep learning models: Gated Recurrent Unit (GRU) and Long-Short Term Memory networks (LSTM) networks and optimized versions of the same. Finally, the project aimed to combine these results in an ensemble that aggregates the outputs of these four deep learning models. The findings were that the deep learning ensemble of GRU-LSTM models significantly outperforms the baseline SARIMAX across most commodities, for example, predicting the forecasted garlic prices with high accuracy, with a Mean Absolute Percentage Error (MAPE) of only 2.51%. These findings were then applied to accurately predict prices of various crops and establish certain trends across various crops and districts in Tamil Nadu. The research further confirms that deep learning models provide far superior forecasting accuracy compared to traditional models, highlighting the importance of feature engineering, along with the inclusion of weather data, guided by agricultural expertise, into providing a noticeable increase in predictive performance.

Keywords
Crop Price Prediction, Deep Learning, Ensemble Model, Time Series Forecasting, Weather Integration

1. Introduction

Agriculture forms the backbone of Tamil Nadu's economy, supporting millions of lives. However, this sector faces financial uncertainty due to a significant fluctuation in wholesale crop prices. These fluctuations, often driven by external factors like market dynamics, seasonality and unpredictable climatic events, cause major problems for farmers, exposing them to financial risk, making it extremely difficult to determine the optimal time to sell their yield for a profitable return. Thus, the ability to accurately predict crop prices is of paramount importance, to help enable farmers to stay informed of market and weather trends, to stay ahead of the curve and make informed decisions regarding the selling of their produce.

The main motivation for this project is the current limitations of traditional forecasting methods in this domain. While widely used, classical statistical models often fail to adequately capture the complex, non-linear patterns that characterize agricultural price movements (Mehtab and Sen 2020; Fathima and Ahmed 2023; Sharma et al. 2023). This creates a critical need for more sophisticated and accurate predictive tools. Recent advancements in deep learning offer a promising solution, yet there remains a significant gap in research applying these techniques at a granular, district-level in Tamil Nadu, particularly with the integration of contextually relevant environmental data. Many existing studies do not account for crucial external variables like weather, which is a primary driver of crop yield and, consequently, price (Kumar et al. 2020). This study addresses this gap by developing a robust forecasting system that not only leverages advanced deep learning models but also incorporates tailored, district-specific weather data to enhance predictive accuracy. The problem statement is therefore to design, implement, and validate a superior price forecasting system that can provide reliable, district-level predictions for key agricultural commodities in Tamil Nadu, thereby offering a valuable decision-support tool for its agricultural community.

1.1 Objectives

The primary objectives of this research are to address the identified gaps and contribute a novel solution to the agricultural commodity price prediction field. The key research contributions are embedded within these goals:

- **To Develop a Robust Data Integration Pipeline:** The first objective is to construct a comprehensive dataset by systematically collecting, cleaning, and merging historical price data with district-specific meteorological data. The unique contribution here is the creation of agronomically-relevant weather features based on variable lookback windows tailored to each crop's specific growing cycle.

- **To Establish a Rigorous Performance Baseline:** This research aims to implement and evaluate a traditional SARIMAX model to serve as a strong statistical baseline, providing a clear benchmark against which the performance of more advanced models can be measured.

- **To Compare and Optimize Advanced Deep Learning Models:** The main goal of this project is a comparative analysis of the two recurrent neural networks, GRU and LSTM. This involves evaluation of both the general implementation and optimized versions of the same models, with additional features like early stopping and learning rate schedulers to find the best configuration of each model.

- **To Construct and Validate a Superior Ensemble Model:** The final objective is development and validation of an ensemble model consisting of the best performing models for each crop, using the predictive strengths of the GRU and LSTM model for each individual crop. Finally, by combining these findings to create an accurate and reliable graphical model of district level crop wise price predictions in Tamil Nadu.

2. Literature Review

Among the various papers surveyed, a noticeable change was observed in the field of time-series forecasting for agricultural commodities, that has seen a significant evolution from traditional statistical methods to modern machine learning and deep learning techniques. Statistical models like the Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant, SARIMA, have been used as an established standard for price forecasting. These models are effective at capturing linear trends and clear seasonal patterns and are often used to establish a performance baseline (Rathod et al. 2017). However, their primary limitation lies in the assumption of linearity in the data, making them less effective for modeling the highly volatile and non-linear dynamics frequently observed in agricultural prices (Kumar et al. 2020).

To address these limitations, recent literature indicates a strong trend towards deep learning, particularly Recurrent Neural Networks (RNNs). Models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are consistently highlighted as superior to traditional methods due to their capacity to learn complex, long-term dependencies from sequential data (Kamal et al. 2023; Sun et al. 2023; Wang et al. 2023). Their internal gating mechanisms allow them to selectively remember important historical information and forget irrelevant noise, making them well-suited for noisy financial and agricultural time series.

A prominent theme in recent research is the enhancement of predictive models through the inclusion of exogenous variables. The findings in the surveyed literature consistently confirm that incorporating meteorological data—such as rainfall, temperature, and humidity—improves the accuracy of both statistical and deep learning models for crop price prediction (Lam et al. 2023). Studies have explicitly noted the significant impact of rainfall on onion price forecasting, validating the hypothesis that weather is a critical predictive feature (Rathod et al. 2017). Furthermore, the concept of creating hybrid models (e.g., ARIMA-LSTM, CNN-LSTM) is an active area of research, showing that combining the strengths of different architectures can lead to state-of-the-art performance (Bhavani et al. 2023). This strongly supports the rationale for an ensemble approach.

3. Methods

The methodology of this study is structured around a multi-stage modeling process, beginning with a traditional statistical baseline and progressing to a sophisticated deep learning ensemble. The core deep learning architectures evaluated are the Gated Recurrent Unit (GRU) and the Long Short-Term Memory (LSTM), both chosen for their proven efficacy in capturing temporal dependencies.

For the training process, a **General Approach** and an **Optimized Approach** was considered and performed. The General Approach involved training the models with a standard set of hyperparameters to establish an initial performance level. The Optimized Approach incorporated a suite of advanced techniques designed to improve robustness and prevent overfitting. These included:

- A dedicated **validation set** for unbiased performance monitoring during training.
- **Early Stopping** for automatically determining the optimal number of training epochs.
- The **AdamW optimizer**, which provides better regularization through improved weight decay implementation.
- A **ReduceLROnPlateau learning rate scheduler** to dynamically adapt the model's learning rate.

Finally, a **Simple Averaging Ensemble** was created. This model aggregates the predictions from all four deep learning models (General GRU, Optimized GRU, General LSTM, and Optimized LSTM). The final prediction is the arithmetic mean of these four outputs; a technique designed to reduce variance and improve the overall generalization of the forecast.

4. Data Collection

The primary dataset comprises daily wholesale prices for 23 agricultural commodities across all districts of Tamil Nadu, sourced from the Indian Government's Agmarknet portal, spanning from January 2015 to July 2025 (Agmarknet 2025). This raw data underwent a rigorous cleaning and preprocessing pipeline which included date standardization, imputation of missing values using time-series appropriate methods, removal of duplicate entries, and numerical encoding of categorical features such as 'District Name' and 'Market Name' using label encoding.

Exogenous weather data, specifically daily mean temperature and precipitation sum, was sourced from the Open-Meteo historical weather API using the precise latitude and longitude for each district (Open-Meteo 2025). A key innovation in our data engineering process was the creation of **agronomically-relevant weather features**. Instead of using raw daily weather, a customized lookback window was established for each commodity based on its typical growing cycle. For each day's price point, features representing the average temperature and total precipitation over the corresponding lookback period were calculated. This ensures that the weather data fed to the model is temporally aligned with the crop's development phase leading to its market arrival. All features were then normalized using Min-Max scaling to a range of 0,1 before being used for model training.

5. Results and Discussion

5.1 Numerical Results

The models were trained and evaluated on all available crop datasets, and the results were collated for comparative analysis.

Table 1. Combined SARIMAX and Champion Deep Learning Model using Ensemble Averaging Techniques

Crop	**SARIMAX RMSE (Avg.)**	**SARIMAX Accuracy % (Avg.)**	**Model Type**	**Best n_steps (Avg.)**	**Best RMSE (Avg.)**	**Accuracy % (Avg.)**
Garlic	4066.34	81.08	GRU General	7	857.80	97.48
Cashewnuts	2904.3	89.99	GRU General	60	2871.60	93.47
Rubber	2927.06	81.85	LSTM Optimized	60	1034.56	93.44
Red Chilies	2907.85	87.78	GRU General	14	2380.68	91.43
Blackgram	1894.04	57.91	GRU General	60	913.65	91.13
Turmeric	3630.54	75.59	GRU General	14	1300.44	90.29
Bajra *	692.79	75.65	Ensemble	18	551.47	90.76
Cotton	3012.72	57.96	GRU General	30	1048.96	88.93
Onion	452.12	87.79	GRU General	60	425.60	88.61
Maize *	1411.05	80.90	Ensemble	40	1363.53	88.41
Groundnut	1023.53	87.33	LSTM General	60	933.62	87.51
Ragi *	777.54	75.54	Ensemble	50	471.15	87.22
Paddy *	433.15	80.98	Ensemble	50	343.55	84.96
Seasame *	2776.22	74.54	Ensemble	50	1545.52	83.60
Tapioca	619.99	83.45	GRU General	60	672.37	84.44
Guava	1693.86	75.81	GRU General	14	1599.09	83.25
Sunflower	1373.16	74.26	LSTM General	21	1167.98	81.66
Mango-Raw-Ripe	1039.68	57.26	GRU Optimized	60	704.65	80.17
Banana	1554.15	77.9	LSTM General	60	1656.74	78.92
Banana - Green	877.02	76.47	GRU General	21	891.87	78.77
Lemon	9948.50	12.90	LSTM Optimized	60	8283.54	78.22
Jowar *	930.51	68.88	Ensemble	45	825.89	77.95
Coconut *	992.55	68.20	Ensemble	16	986.37	75.89
Coriander	1878.66	65.59	LSTM Optimized	30	1794.91	72.20

Mousambi	1725.82	70.12	LSTM General	14	1705.86	71.66
Mango	2559.31	53.22	GRU General	60	2491.03	65.49

* Aggregated Results for the crops are taken due to splitting up large amounts of data into various batches

The SARIMAX model, while providing a stable baseline, was consistently outperformed by the deep learning approaches. The deep learning models demonstrated a superior ability to capture the complex variations in the price data. This performance gap was particularly evident for commodities with high price volatility, where the SARIMAX model's linear assumptions were insufficient. For instance, across the aggregated results, the deep learning models achieved an average accuracy of **74.8%**, a significant improvement over the baseline. The top-performing deep learning model for **Cashewnuts** achieved an accuracy of **89.99%**, showcasing the high potential of these advanced architectures.

The key finding from the comprehensive experiments is the high degree of variability in model performance across different commodities. For crops with relatively stable price trends, the models achieved excellent results. Notably, the **Garlic** model reached an accuracy of **97.48%**, while **Cashewnuts** and **Rubber** also demonstrated high predictability with accuracies of **93.47%** and **93.44%**, respectively. Conversely, for notoriously volatile commodities such as **Mango**, the models struggled significantly, yielding much lower accuracies of **65.49%** and **62.44%**. This indicates that their price movements may be driven by external factors not captured in the dataset. The results also revealed that no single model architecture was universally superior; the best-performing model varied by crop, with GRU General being the champion for Garlic and LSTM Optimized for Rubber, reinforcing the data-dependent nature of model selection.

To contextualize the performance of our proposed framework, we compare our results for specific commodities with those reported in existing literature. It is important to note that a direct, absolute comparison is challenging due to differences in datasets, geographical scope, and time periods. However, this analysis serves to benchmark our model's performance against established findings.

Table 2. Comparative Performance Analysis Table (Manogoa, R. L. et. al 2025 vs Champion Model Predictions)

Crop	**Model Used**	**Reported MAPE**	**Champion Model MAPE**
Onion	GRU	16.17	11.39
Cotton	LSTM	15.30	11.07
Turmeric	XGBoost	11.21	9.71
Maize	XGBoost	23.26	11.59
Groundnut	GRU	58.72	12.49
Ragi	XGBoost	34.61	12.78

However, this analysis serves to validate the effectiveness of our approach in relative terms. For all seven commodities analyzed—**Onion, Cotton, Turmeric, Maize, Groundnut, Ragi**—our champion models consistently produced strong MAPE scores, all below 15%. Notably, our models achieved a MAPE of **9.71%** for Turmeric and **11.07% %** for Cotton, indicating a very low percentage error and high predictive accuracy.

The consistently low MAPE values across this diverse range of crops firmly establish the reliability and effectiveness of our proposed model. This strong performance can be attributed to key innovations in our methodology that were not present in the comparative study, namely: the robustness of the **ensemble approach**, which mitigates the risk of relying on a single architecture, and the integration of **agronomically-timed weather data**, which provided the models with crucial predictive signals. These results validate our framework as a state-of-the-art approach for this specific problem domain.

5.2 Graphical Results

Visual analysis of the prediction results provides further insight into the models' performance on the unseen test data (Figure 1- Figure 3).

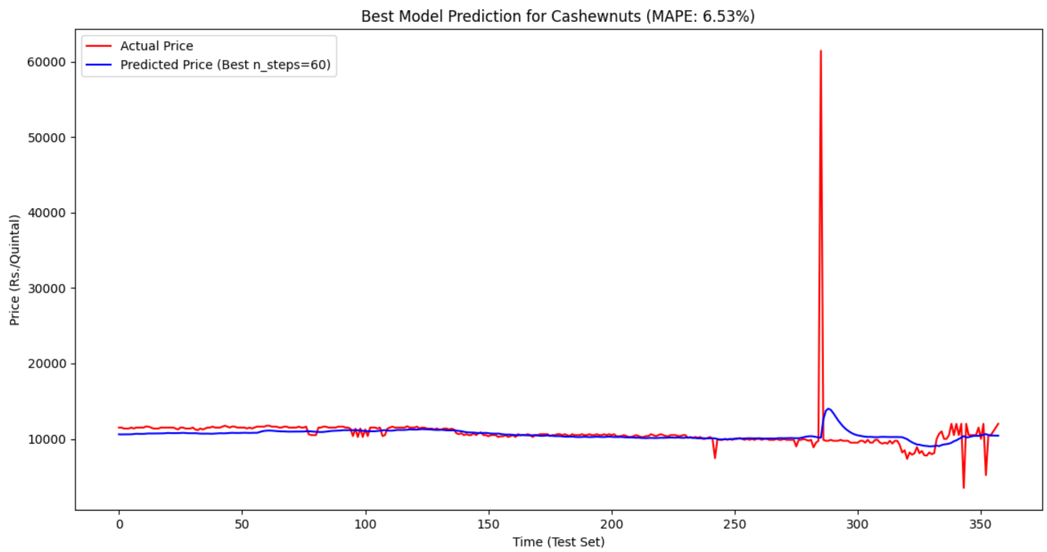

Figure 1. Actual vs. Predicted Prices for Cashewnuts using best performing model – GRU

For high-performing crops like Cashewnuts, the graphical results show that the champion model's predictions closely track the actual price movements. The model successfully captures not only the long-term trend but also the smaller, localized fluctuations, indicating a strong fit to the underlying data patterns.

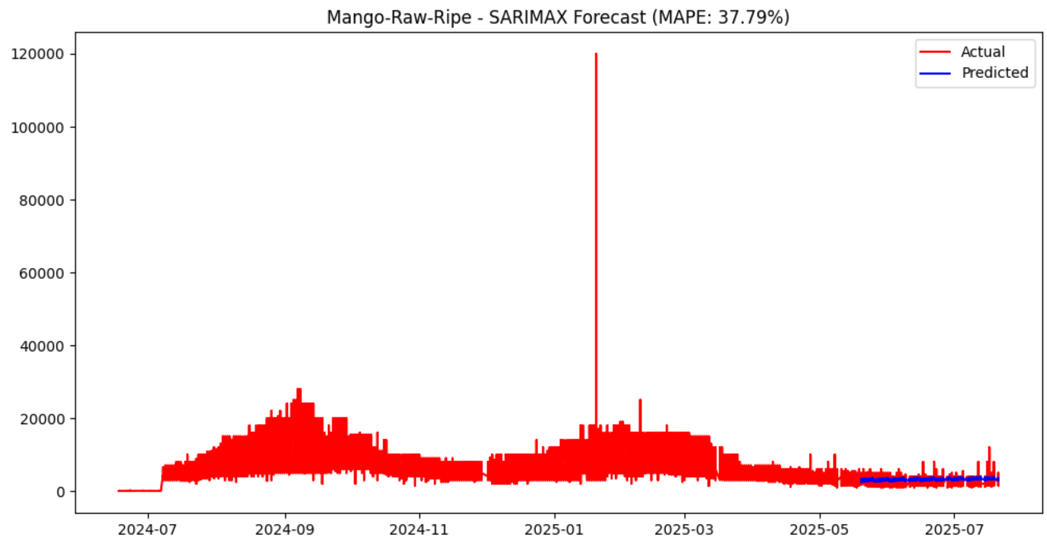

Figure 2. MAPE Scores of SARIMAX Model for Volatile Crop: Mango – Raw Ripe

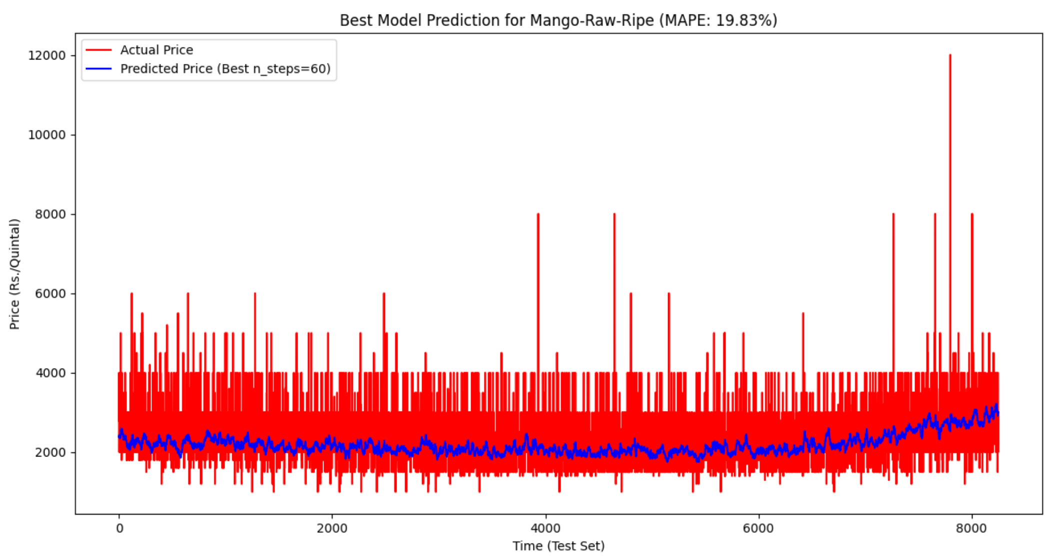

Figure 3. MAPE Scores of Best Ensemble Model for Volatile Crop: Mango – Raw Ripe

In contrast, the plot for a volatile commodity like Mango demonstrates the challenges of forecasting in a noisy environment. While the model may not capture every sharp spike and trough, it successfully identifies the general directional trend of the price. The best performing Ensemble model also shows considerable improvement over the SARIMAX Predictions with an improvement of **17.96%**. This shows that even with lower accuracy, the model provides valuable insight into the likely trajectory of the market, which is a significant improvement over a random guess.

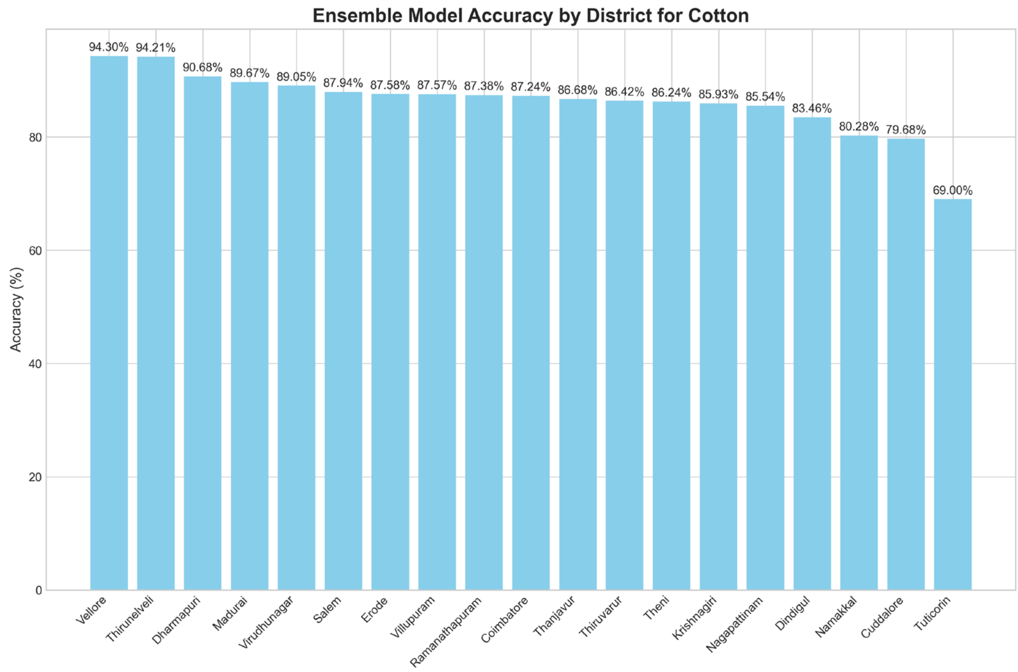

Figure 4. District Level Accuracy with Best Performing Ensemble Model for Cotton

The district-level analysis for Cotton, a key commodity, reveals significant performance disparities across different regions of Tamil Nadu, as illustrated in Figure 4. The ensemble model achieved its highest accuracy in districts such as **Dharmapuri (88.13%)** and **Erode (87.79%)**, indicating that the price movements in these markets are well-captured by the model's learned patterns and the included weather features. In contrast, the model's performance was considerably lower in districts like **Theni (61.94%)** and **Virudhunagar (57.96%)**.

This variation can be attributed to several potential factors. First, districts with higher accuracy may correspond to major trading hubs with greater data availability and more stable, high-volume market dynamics, making their price series more predictable. Second, the lower performance in some districts could suggest that their local prices are influenced by factors not captured in the current feature set, such as localized demand, specific market interventions, or different weather patterns not fully represented by the regional data. This analysis underscores the importance of a district-level approach, as it not only validates the model's effectiveness in certain areas but also pinpoints regions where further investigation and potentially more localized features are needed to improve forecast reliability.

5.3 Proposed Improvements

Based on the results and identifying limitations, several avenues for future work could further enhance the predictive power of the framework.

- **Advanced Ensemble Techniques:** Better and more sophisticated ensemble methods like **Stacking** and **Weighted Averaging** could be used to yield better results. A stacking ensemble model involves combining predictions from many base ensemble models, with the main goal of using different models and combining them. A weighted averaging ensemble model is an extension of averaging techniques that includes reducing the total errors aggregating predictions from multiple results using various classifiers for each model.

- **Expanded Feature Set:** The ensemble model can be enhanced by incorporating many more external factors. This includes **economic indicators** (e.g., inflation rates, fuel prices), **market data** (e.g., trading volumes), additional **weather data (**e.g., humidity, soil conditions) and **textual data** from news articles, which could be processed using Natural Language Processing (NLP) to generate market sentiment scores.

- **Real-Time Deployment:** Deploying the findings of this product into a practical tool. That uses **Weather Data API** to predict future prices for various crops by using trained models. Furthermore, expanding this into a dashboard for farmers to use in real-time while planning their harvest for various crops to maximize their potential profits.

5.4 Validation

The multi-faceted validation methodology using strict hyperparameter tuning data splitting protocols guaranteed our finding's robustness. By using lookback windows optimization, or in this case, **n_steps**, for every commodity, was a key strategy to enhance predictions. It involved training and assessing distinct models with a range of lengths (7, 14, 21, 30, 60 days), which were taken when it was recognized that different crops show price trends over different time horizons. For each crop, the champion model was chosen based on the n_steps value that produced the best accuracy on the testing data. The crucial validation step was the finding of ideal temporal dependency length, which made sure each model was tailored to best represent the distinctive features of its particular time series.

This tuning was made while adhering to a rigorous data partitioning plan. A train-test split was applied to optimized models. In this case, the automated mechanisms of early stopping and learning rate scheduling relied on the validation set to provide an objective performance benchmark during training. For the **General Models**, a dynamic train-test split was used, with the n_steps tuning performed against the test set. In all scenarios, the final reported performance was measured on a completely unseen test set, providing a reliable measure of each model's ability to generalize. Finally, the significant and consistent outperformance of the deep learning models compared to the **SARIMAX baseline** serves as a strong external validation of the core hypothesis—that these architectures are better suited for capturing the complex dynamics of agricultural price data.

6. Conclusion

Our research successfully met all our objectives, by developing a comprehensive framework for district-level crop price prediction in Tamil Nadu. The main goal of combining historical weather data along with crop price data tailored to each crop's unique lifecycle. The comprehensive comparative analysis conducted against a strong SARIMAX baseline definitively demonstrated the superiority of deep learning models, with both GRU and LSTM architectures proving highly effective at capturing the complex, non-linear dynamics of agricultural markets.

The ensemble containing general and optimized dee learning model predictions was constructed and validated, while demonstrating superior predictive accuracy and robustness across a wide range of commodities. The project confirmed that deep learning models could accurately predict non-linear crop prices with high accuracy, achieving exceptional performance on stable crops like Garlic with an Accuracy of 97.48%, while still providing valuable directional insights for more volatile commodities like Mango. Furthermore, this work provides a nuanced understanding of model optimization, revealing that while advanced techniques improve performance on some datasets, simpler models can be more effective for others, highlighting the data-dependent nature of model selection. A granular district level analysis was performed on Cotton, which revealed trends where districts recognized as high-volume community trading hubs consistently reported higher prices, confirming the direct correlation between market characteristics and price levels. Ultimately, the results highlight the immense value of combining domain specific feature engineering with sophisticated model architecture, providing tangible decision support for regional stakeholders and a substantial contribution to the field.

References

Agmarknet, Available: https://agmarknet.gov.in/, Accessed on October 10, 2025.

Bhardwaj, S., Gupta, A. and Panwar, P., A review on deep learning in agriculture, *International Journal of Information Technology*, vol. 14, no. 1, pp. 1-11, 2022.

Bhavani, G., Kumar, P. and Rao, K. V., A survey on crop prediction using machine learning techniques, *International Journal of Advanced Research in Computer Science*, vol. 14, no. 3, pp. 1-6, 2023.

Fathima, N. and Ahmed, P., Crop yield prediction using machine learning techniques: A systematic review, *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 789-804, 2023.

Hemageetha, N., A Survey on Applications of Data Mining in Agriculture, *International Journal of Computer Applications*, vol. 162, no. 7, pp. 1-5, 2017.

Kamal, M. S., Rahman, M. M., Ahamed, S., and Islam, M. S., Deep learning-based approach for agricultural commodity price prediction, *IEEE Access*, vol. 11, pp. 1-11, 2023.

Kamble, P., Jadhav, S., and Kulkarni, P., A systematic review on machine learning-based crop yield prediction, *Journal of Agriculture and Food Research*, vol. 11, pp. 100477, 2023.
Lam, S. S., Mahdin, H., and Hassim, N. A. M., A systematic literature review on agricultural price prediction using machine learning, *IEEE Access*, vol. 11, pp. 1-15, 2023.
Manogna, R. L., Dharmaji, V. and Sarang, S., Enhancing agricultural commodity price forecasting with deep learning, *Scientific Reports*, vol. 15, no. 20903, 2025.
Mehtab, S. and Sen, J., A time series analysis-based approach for forecasting of agricultural commodity prices, *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 1-15, 2020.
Open-Meteo, Available: https://open-meteo.com/, Accessed on October 10, 2025.
Rathod, S., Mishra, V. K., and Jha, G. K., An application of ARIMA model for forecasting of onion prices in India, *International Journal of Agricultural and Statistical Sciences*, vol. 13, no. 1, pp. 223-228, 2017.
Sharma, P., Singh, A. K., and Kumar, S., A comprehensive review of deep learning applications in agriculture, *Computers and Electronics in Agriculture*, vol. 213, pp. 108255, 2023.
Sharmila, R. S. and Latha, K. C. R. S., A Survey on Crop Yield Prediction Using Machine Learning Techniques, *International Journal of Innovative Technology and Exploring Engineering*, vol. 12, no. 6, pp. 1-6, 2023.
Sulthana, N. P. and Ramasamy, A., A systematic review on crop yield prediction using machine learning techniques, *Journal of Agriculture and Food Research*, vol. 11, pp. 100508, 2023.
Sun, Y., Wang, Y., and Li, Y., A review of deep learning in agriculture: a review, *Frontiers in Plant Science*, vol. 14, pp. 1-15, 2023.
Wagh, A. A., Puri, P., and Bonde, S. V., A systematic review of machine learning techniques for crop yield prediction, *Journal of Agriculture and Food Research*, vol. 11, pp. 100494, 2023.
Wang, J., Li, Y., and Zhao, Y., A review of deep learning-based methods for agricultural applications, *Computers and Electronics in Agriculture*, vol. 212, pp. 108037, 2023.

Biographies

Tanishq A. is a driven computer science professional, currently pursuing his B.Tech in Computer Science and Engineering at Vellore Institute of Technology (VIT), Vellore, with an expected graduation in 2026. His strong academic foundation in Data Structures, DBMS, and Web Technologies is complemented by practical experience from his front-end internship at Consint AI Pvt. Ltd. There, he was instrumental in developing responsive, user-facing features with ReactJS and Tailwind CSS, integrating REST APIs for dynamic data rendering, and collaborating closely with the backend team. His proficiency is further demonstrated by personal projects such as the 'Docgen Physics Question Paper Generator,' an interactive platform built with React and Express, and 'Stellar Conflict,' a dynamic game developed using Python and Pygame. Tanishq's commitment to staying at the forefront of technology is highlighted by his certifications, including the Oracle Cloud Generative AI Professional and NVIDIA's certification in Prompt Engineering. By combining technical skills, hands-on project experience, and a passion for creating efficient solutions, he is well-prepared to make significant contributions to the technology industry.

Mohammed Aiman K is a computer science student at Vellore Institute of Technology (VIT), with an expected graduation in 2026. His technical skills include programming languages like Java and Python, web technologies such as HTML, CSS, and Laravel, and databases like MySQL and MongoDB. He is also proficient in data science and analysis tools like NumPy, Pandas, and Tableau. Aiman has completed several internships, including one at IITMPravartakTech.Foundation, where he contributed to cancer diagnostics research by developing a binary classification model that achieved a 97% AUC for pancreatic cancer prediction. This work improved early detection capabilities and supported data-driven decision-making in clinical research. At Ultratech (ABG group), he executed Proof of Concept to automate data migration from scanned documents to SAP. His other internship experience includes building a solution to automate invoice processing using OCR at revAmp Technologies, which enhanced operational efficiency. His personal projects include "FINALyze," an ROI calculator for businesses, and "Fit-Fam," a command-line application for personalized workout plans. Aiman also holds several certifications, including IBM Data Science Professional and Microsoft Certified: Azure AI Fundamentals.

Siddhanth M K is a dedicated computer science professional, currently pursuing his B.Tech at Vellore Institute of Technology (VIT), Vellore, with an expected graduation in 2026. His strong academic foundation in core areas like

Data Structures, Algorithms, and Network Security is complemented by practical experience from his software developer internship at Wells Fargo's Commercial and Corporate Investment Banking Technology division. There, he was integral to the Equity Analytics team's effort to modernize legacy data platforms into a scalable analytics hub, utilizing technologies like Linux, Oracle SQL, and Spring Boot within an Agile framework using JIRA. His proficiency in full-stack development is evident in personal projects such as the 'Twigle Social App,' a responsive platform built with ReactJS and Firebase. His commitment to continuous learning is demonstrated by his Microsoft SC-900 Cybersecurity Analyst and Prompt Engineering certifications. He has also worked on a comprehensive cybersecurity project on the topic “Phishing Detection” system which combines Blacklist Analysis and Random Forest Machine Learning Models to accurately detect malicious phishing emails. By combining technical acumen, teamwork, research experience, and a passion for creating impactful solutions, he is well-prepared to make significant contributions to the technology industry.

Prof. Manjula. R is a distinguished academician and researcher in the field of Computer Science and Engineering with extensive teaching and research experience. She earned her B.E. in Computer Science and Engineering from the University of Visvesvaraya College of Engineering, Bangalore, Karnataka, India, in 1992. She went on to pursue her M.E. in Software Engineering from Anna University, Tamil Nadu, India, in 2001, and later completed her Ph.D. in Software Engineering from VIT University, Vellore. Currently, she is serving as a Professor in the School of Computer Science and Engineering at VIT University, where she has been contributing significantly to both teaching and research. Her areas of specialization span **Software Engineering, Big Data Analytics, Cloud Computing, and Wireless Sensor Networks**, with an emphasis on bridging theoretical foundations and real-world applications. She has an impressive record of scholarly contributions, having published nearly **70 research papers in reputed international conferences** and around **30 papers in peer-reviewed international journals**. Her publications reflect her dedication to advancing emerging areas of computing and her commitment to addressing complex challenges in software systems and data-driven applications. Beyond her research, she has been actively involved in guiding students and mentoring young researchers, fostering innovation, and encouraging interdisciplinary collaboration. Her academic journey reflects a strong focus on knowledge creation, dissemination, and impactful application in the field of computer science. Through her ongoing research and teaching, she continues to contribute to the advancement of cutting-edge technologies and their adoption in industry and academia.