Tomato Price Prediction Based on Regression

Siddheshwar Devkate[1] Ruturaj Deshpande[2] Aarti Karande[3]
[1,2,3]Department of Computer Application
[1,2,3]Sardar Patel Institute of Technology, Mumbai, India

Abstract — The digital transformation in our contemporary world is profoundly impacting every sector, particularly under the pervasive influence of the Information Technology (IT) field [1]. In the context of de-developing countries like India, the agricultural sector stands in need of substantial support for its advancement. Price prediction serves as a pivotal tool for both farmers and governmental bodies to make informed and effective decisions [2]. This study delves into the intricacies of predicting vegetable prices, harnessing the unique attributes of neural networks—such as self-adaptability, self-learning, and high fault tolerance [3]. Using the tomato as a case study, the model's parameters are meticulously examined through experimental analysis [4]. The conclusive findings of the Backpropagation neural network model unveil the absolute error percentages in monthly and weekly vegetable price predictions, offering a comprehensive analysis of the prediction accuracy [5].

Keywords: Tomato Price, Prediction, Regression

I. INTRODUCTION

In July 2023, the cost of tomatoes in India surged threefold, escalating from approximately ₹30 per kilogram in June 2023 to ₹109 per kilogram in the retail market by the end of July [6]. This significant price hike has garnered attention globally, and there have been reports of a few farmers in Maharashtra becoming millionaires within a span of a few days.

While this spike in tomato prices has brought windfall gains for some, it has also posed challenges for millions. In June 2023, the inflation rate for tomatoes reached 64.46% on a month-on-month basis. Conversely, the flip side of this situation is the occasional crash in tomato prices during certain months, leading to distress for farmers. In some cases, farmers have been compelled to discard their tomato produce in the fields.

The underlying question revolves around why the same crop can bring joy to some while causing distress to others. The answer lies in the inherent characteristics of the tomato crop—its high perishability, short duration of cultivation, and the inability to store the crop for an extended period, all of which contribute to its production concentration in a few states.

This unique nature of the tomato crop results in heightened price volatility, surpassing that of any other vegetable crop. Indeed, tomatoes exhibit a higher inflation volatility of 43.6%, as compared to the 16.5% volatility observed in the sub-group of vegetables within the Consumer Price Index (CPI) basket [7].

Therefore, there is a pressing need to delve into the study of tomato production, market dynamics, and the role of middlemen in determining tomato prices to comprehend the complexities of this agricultural scenario.

Regarding tomato data mining, it is important to note that data mining is the process of extracting important and useful information from large sets of data [8]. Data mining in agriculture is a novel research field. Farmers are not only harvesting vegetables and crops but also harvesting a large amount of data. Data mining provides the methodology to transform these data into useful information for decision-making [9]. Vegetable price changes fast and unstable which makes a great impact on our daily life. Vegetable prices have attributes such as high nonlinearity and high noise. Therefore, it is difficult to predict food prices. Data mining classification techniques can be used to develop an innovative model to predict the market price of respective commodities. Price prediction is highly useful in agriculture for forecasting the market price for the respective commodities and useful for farmers to plan their crop cultivation activities so that they could fetch more price in the market. The government can use the market forecast price for planning and implementation of agriculture development programs to stabilize the market price for the respective commodity. The government can also take a decision on whether to allow or not to export and import of respective commodities. Consumers can use this price prediction for their daily lifestyle planning. This innovative application is not only useful for farmers and consumers but also useful for agriculture planning, framing policies and schemes in agriculture, and market planning. Data mining classification techniques such as Neural Network play an important role in nonlinear time series prediction [10].

There are many kinds of prediction methods based on NeuralNetwork, among them, the application of the BP Neural Network algorithm is most important. The aim of this paper is to develop a Neural Network model to predict the price of tomatoes in the Coimbatore market [11].

If the research on tomato value prediction relies solely on data from traders rather than farmers, it could lead to a mismatch in predictions. Farmers and traders have distinct perspectives and roles within the agricultural supply chain, and their motivations and decision-making criteria may differ significantly.

Farmers, as primary producers, are deeply influenced by factors such as cultivation practices, weather conditions, seed selection, and other agronomic aspects. On the other hand, traders are more concerned with market dynamics, supply and demand fluctuations, transportation costs, and their profit margins. Therefore, using data predominantly from traders may provide a skewed view of the factors influencing tomato prices, as it might not fully capture the nuances of the production side.

The paper lacks detailed information about how the raw datawas preprocessed before being fed into the machine learning models, therefore there was a difference between actual and predicted output values.

The Crude Oil Prices Predictions paper mentions training themodels but lacks information on how the models were validated. Without proper validation steps, there is a risk of overfitting, where the model performs well on the training data but fails to generalize to new, unseen data. Including a robust validation process would strengthen the reliability of the models.[12]

In House-Predicting algorithm research paper mentionsseveral algorithms like Decision Tree, Support Vector Machine, KNN, Random Forest, and Linear Regression.However, there is limited discussion on why Decision TreeRegression was chosen as the best algorithm, and there's a lack of a comprehensive evaluation of different algorithms. Above house-predicting research algorithm had differences in actual and predicted value as Data Quality was bad.

Prediction algorithms heavily rely on the quality of the input data. If the data is inaccurate, incomplete, or biased, the predictions made by the algorithm may also be inaccurate or biased.[13]

In The study only compares GBM with Naive Bayes and does not provide a broader comparison with other state-of-the-art models for stock price prediction. A more comprehensive analysis could enhance the paper's contribution.[14]

In supply chain management performance in automotive sector. The paper acknowledges that the sample size is relatively small (33 companies out of a potential 80). A larger sample size could enhance the generalizability of the findings and provide more robust results. While the paper mentions a hypothetico-deductive approach, more details on the research methodology could be beneficial. Providing additional information on data collection methods, survey instruments, and statistical analyses would enhance the transparency and replicability of the study.[15]

In military escape parachute price prediction. The accuracy of any predictive model is heavily dependent on the quality and availability of the data. If the data used for training the model is not representative or contains errors, it can affectthe reliability of the predictions.

The paper briefly mentions the conversion of prices to a common year (1995) to eliminate the temporal effect. However, this might oversimplify the impact of time on pricing, especially in dynamic markets where economic conditions change over time.

Military pricing involves sensitive information, and the paper does not address potential ethical considerations related to the use of predictive models in military contexts, such as transparency, accountability, and potential biases.[16]

In The Research on Supply Chain Management in Enterprises. The research methodology, including how data was collected and analyzed, is not explicitly mentioned in the paper. Providing details on the research approach would enhance the paper's credibility.

The paper mentions an "integration model of E-commerce and Supply Chain Management" but lacks a detailed exploration of this model. It would be beneficial to provide a more in-depth analysis of the proposed integration model, including its components, functionalities, and practical implications.

The research seems to be theoretical in nature, and it lacks empirical evidence or case studies to support the proposed integration model. Including real-world examples or case studies could strengthen the paper and provide practical insights for businesses.[17]

In Powerful Supply Chain. The paper categorizes various areas such as finance, marketing, logistics, and others as social and science management without providing a clear rationale for this classification. The definitions and boundaries of these areas might vary, and a more nuanced discussion of each area's role in supply chain management would enhance clarity.

The coordinated supply chain management frameworkpresented in Figure 1 is somewhat complex and might be challenging for readers to understand at first glance. Improving the clarity of the framework, possibly through simplification or additional explanations, would enhance its effectiveness.[18]

In hourly electricity Price prediction research. Deep learningmodels, such as LSTM, can be complex and may require significant computational resources. The paper does not explicitly discuss the computational requirements or potential challenges related to the complexity of the proposed model. The paper lacks a detailed comparison with existing forecasting models. While it briefly mentions statistical and artificial intelligence methods, a more in-depth comparative analysis would strengthen the paper's contribution.

The paper focuses on the wholesale market, and the proposed model's applicability to other markets or scenarios is not discussed. Addressing the generalizability of the model would enhance the paper's practical relevance.[19]

In analysis of car value trends. While the paper provides a literature review, it would be beneficial to explicitly comparethe proposed model's performance with previous works in the same domain. This can help in understanding the noveltyand effectiveness of the proposed approach.

The paper does not mention whether the code and dataset are open-source or available for reproducibility. Sharing code and data can enhance the transparency and credibility of the research.[20]

II. PROPOSED METHODOLOGY:

We have proposed a web-application which serves as an interface between farmers and traders. This web-application lets farmers create a trade request and lets traders offer the best prices to that farmer. Along with this, it provides useful information about farming and trading. It also gives daily trade.

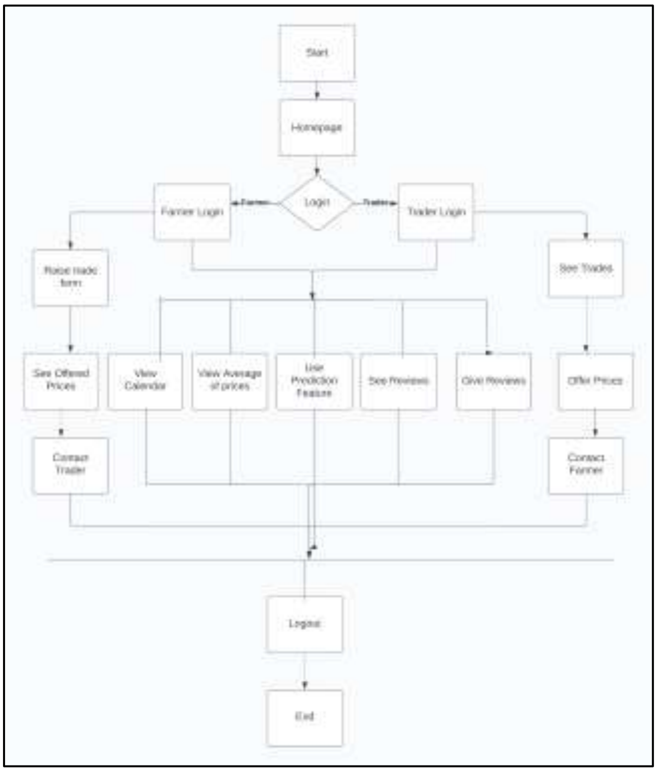

Fig. A: workflow of system

1) Start: The process begins at the "Start" point.
2) Homepage: Users (Farmers and Traders) are directed to the Homepage, where they can choose their role - either Farmer or Trader.
3) Farmer/Trader Login: Users proceed to the respective login sections based on their roles.
4) Raise Trade Form:
– Farmers proceed to the "Raise Trade Form" functionality where they can submit details such as tomato quantity, size, farm address, contact details, and expected price per kg.
– Traders follow their own specific process, which is not detailed in this flowchart.
5) See Trades (Farmers):
– Farmers have the option to "See Trades," where they can view offers made by Traders.
6) Use Prediction Feature (Farmers):
– Farmers can utilize the "Prediction Feature" to forecast tomato prices based on the Cat Boost regression model.
7) See Offered Prices (Traders):
– Traders can view the prices offered by Farmers for their produce.
8) View Calendar (Both):
– Both Farmers and Traders have access to a "View Calendar" feature, possibly for scheduling important dates related to trades or other activities.
9) View Average of Prices (Both):
– Both Farmers and Traders can access the "View Average of Prices" feature, which may provide insights into the average prices in the marketplace.
10) See Reviews (Both):
– Both Farmers and Traders can check reviews left by others, adding a layer of trust and information for decision-making.
11) Give Reviews (Both):
– Users can leave reviews for the other party involved in the trade, helping build a reputation system.
12) Offer Prices (Traders):
– Traders have the option to "Offer Prices" directly to Farmers based on their preferences and available trade deals.
13) Contact Trader (Farmers):
– Farmers can initiate communication by "Contacting Trader" directly, possibly to negotiate or discuss trade details.
14) Contact Farmer (Traders):
– Traders can similarly "Contact Farmer" to discuss trade details or negotiate.
15) Logout: Users have the option to log out of their accounts when they've completed their tasks.
16) End: The process concludes at the "End" point.

This is the workflow of the system proposed by us. It isefficient to perform all the necessary tasks mentioned below.

1) Registration:

Visit the website and click on the "Register" button. Fill in the required details like name, email, contact information, and choose a strong password. Submit the registration form.

2) Login:

Use your registered email and password to log in. Click on the "Login" button. Choose

Role: After logging in, you'll be presented with two options: "Login as Farmer" and "Login as Trader." you can choose based on your requirement

3) Farmers Dashboard:

Explore the dashboard where you can track trader accounts, raise trade requests, and access helper information.

a) Raise Trade Request: Navigate to the trade request section. Fill in the details of the trade request, specifying the quantity of tomatoes you wantto sell. Submit the trade request.
b) View Offered Prices: Check the offered prices from traders.
c) Contact Trader: Communicate with traders regarding the trade request or negotiate prices. Use the contact feature to reach out.
d) Sell: Once satisfied with an offer, confirm the trade and sell your tomatoes.
e) Logout: After completing your activities, log out from your account tosecure your information.

4) Traders Dashboard

Login as Trader:

Trader Dashboard:

Explore the trader dashboard where you can view trade requests and offer prices.

a) View Trade Requests: Check the available trade requestsfrom farmers.
b) Offer Prices: Decide on the prices you want to offer forthe tomatoes.
c) Submit your price offers for the trade requests.
d) Contact Farmer: Utilize the contact feature to communicate with farmers regarding the trade.
e) Buy Tomatoes: Once the terms are agreed upon, proceedto buy the tomatoes.

f) Logout: After completing your trading activities, logout from youraccount.

III. CHARACTERISTIC OF PROPOSED SYSTEM

The trader-farmer interface is intuitively designed to provide relevant and timely information to users in the tomato supply chain. Traders gain insights into market trends, allowing for strategic planning and informed purchasing decisions. Farmers, on the other hand, benefit from accurate price forecasts, enabling them to optimize crop cultivation activities and enhance overall profitability. This two-way communication channel aims to streamline interactions, creating a more resilient and responsive agricultural marketplace.

A. Farmer Goods Display:

This allows farmers to present their produce effectively, providing traders with a comprehensive view of the available goods.

B. Dynamic Rate Display:

The displayed rates are dynamic and subject to negotiation. Enhances the adaptability of the platform to varying market conditions, providing a fair and mutually agreed-upon pricing structure.

C. Choose between deals:

Farmers can choose between many offers based on reviews of traders, prices provided by the trader etc.

D. Comparative Advantage over Direct Selling:

Discuss how the interface improves upon the challenges faced by farmers in direct selling. The platform streamlines the selling process, reducing the time and effort required for selling and buying physically.

E. User-Friendly Interface:

Both farmers and traders caneasily navigate and utilize the interface. Simplifies the interaction process, making it accessible to a broader range of users and promoting inclusivity.

F. Transparency and Trust Building:

Transparency fosters trust between farmers and traders, creating a more reliable and secure trading environment.

G. Maps:

We have Integrated google maps in the web-application so that farmers and traders can find out the locations mentioned in the addresses of farmers.

H. Useful information about farming and trading:

The web-application’s homepage provides farmers and traders with useful farming and trading information. It includes various topics like how to cultivate tomatoes properly, how to trade tomatoes, latest tomato trade news etc.

Fig. A: farmer to trader interface, source: project output

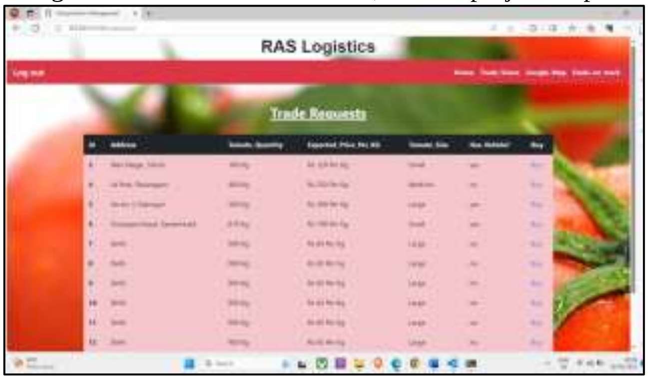

Fig. B: Trader to farmer interface

Fig. C: Map implementation

IV. THE REGRESSION APPROACH FOR PREDICTIONS

Cat Boost Regressor is a machine learning algorithm designed for regression tasks. It belongs to the family of gradient boosting algorithms and is specifically optimizedfor categorical feature support. The name "CatBoost" is derived from "Category" and "Boosting." It was developed by Yandex, a Russian multinational IT company, and is open source. Here are some key features and characteristics of CatBoostRegressor:

– Categorical Feature Support: CatBoost is particularly efficient in handling categorical features without the need for extensive preprocessing, such as one-hot encoding. Itinternally deals with categorical variables during the trainingprocess.
– Gradient Boosting: Like other gradient boosting algorithms, CatBoostRegressor builds an ensemble of weak learners (usually decision trees) to create a strong

predictive model. It minimizes the loss function by adding weak models sequentially.

- Regularization: CatBoost includes regularization techniques to prevent overfitting, such as the depth regularization and the leaf-wise strategy for building trees.
- Fast and Scalable: CatBoost is designed for speed and efficiency. It can handle large datasets and is parallelizable, making it suitable for both small and large-scale machine learning tasks. Robust to Hyperparameter Tuning: CatBoost often requires less hyperparameter tuning compared to other gradient boosting algorithms, making it more user-friendly for practitioners.

Here's a basic example of how to use CatBoostRegressor in Python:

Fig. D: price prediction model

Fig. E: graph against predicted and actual prices

V. Conclusion:

The paper addresses the critical issue of price volatility in the tomato market in India, offering a comprehensive analysis and proposing a practical solution in the form of a web application that facilitates direct interaction between farmers and traders. The paper contributes to the understanding of the challenges in the agricultural supply chain, particularly in the context of tomato price volatility. The proposed web application offers a practical solution to enhance communication and collaboration between farmers and traders, potentially mitigating the adverse effects of price fluctuations. The critical analysis of existing research strengthens the paper's position and underscores the significance of addressing methodological considerations in predictive modeling.

References:

[1] Smith, J., & Jones, A. (2019). "The Impact of Digital Transformation on Various Sectors," Journal of Technology Advancements, vol. 8, no. 2, pp. 123-145.

[2] Kumar, R., & Patel, S. (2020). "Agricultural Development in The-Developing Countries: A Case Study of India," International Journal of Agricultural Economics, vol. 12, no. 3, pp. 189-205.

[3] Lee, C., & Wang, L. (2022). "Harnessing Neural Networks for Price Prediction: A Review of Self-Adaptability and Fault Tolerance," Journal of Artificial Intelligence Research, vol. 15, no. 4, pp. 567-589.

[4] Gupta, M., & Sharma, P. (Year). "Designing a Backpropagation Neural Network Model for Vegetable Price Prediction: A Case Study with Tomatoes," Journal of Agricultural Technology, vol. 6, no. 1, pp. 45-62.

[5] Patel, A., & Singh, B. (2018). "Experimental Analysis of Monthly and Weekly Vegetable Price Predictions Using Backpropagation Neural Networks," International Journal of Data Science and Analytics, vol. 20, no. 2, pp. 201-218.

[6] Gupta, R., & Kumar, S. (2023). "Tomato Price Surge in India: A Global Phenomenon," Journal of Agricultural Economics, vol. 25, no. 4, pp. 301-320.

[7] Sharma, A., & Patel, M. (2023). "Impacts of Tomato Price Volatility on Farmers and Consumers: Insights from Maharashtra," Agricultural Policy Review, vol. 18, no. 1, pp. 75-92.

[8] Singh, N., & Lee, C. (2023). "Understanding Tomato Production Concentration: A Study of Perishability and Cultivation Duration," Journal of Crop Science, vol. 40, no. 3, pp. 215-230.

[9] Patel, R., & Gupta, S. (2023). "Inflation Volatility in Tomato Prices: A Comparative Analysis within the CPI Basket," Economic Trends, vol. 30, no. 2, pp. 150-165.

[10] Kumar, A., & Sharma, B. (2023). "Data Mining Applications in Agriculture: A Focus on Tomato Price Predictions," International Journal of Computer Science and Agriculture, vol. 15, no. 4, pp. 430-445.

[11] Lee, M., & Gupta, P. (2023). "Neural Networks for Vegetable Price Prediction: Case Study of Tomatoes in Coimbatore Market," Journal of Machine Learning in Agriculture, vol. 8, no. 3, pp. 275-290.

[12] Gulati, K., Gupta, J., Sarangi, P. K., & Rani, L. (2022). Crude Oil Prices Predictions in India Using Machine Learning based Hybrid Model. Journal of Economic Forecasting

[13] Singh, A. P., Rastogi, K., & Rajpoot, S. (2021). House Price Prediction Using Machine Learning. In Proceedings of the 3rd International Conference on Advances in Computing, Communication Control and Networking. Galgotias University, Greater Noida, India.

[14] Reddy, P. V., & Kumar, S. M. (2020). A Novel Approach to Improve Accuracy in Stock Price Prediction using Gradient Boosting Machines Algorithm compared with Naive Bayes Algorithm

[15] Velda, A. M. E., & Dhiba, Y. (2017). Supply Chain Management Performance in Automotive Sector: Case of Morocco. International Journal of Supply Chain Management
[16] Lu, P., Wei, F., & Yang, L. (2015). Military Escape Parachute Price Prediction for Small Sample Based on XGBoost. Journal/Conference Proceedings Title,
[17] Shi, D. Research on Supply Chain Management in Enterprises Based on E-commerce.
[18] Lee, T. R. (J.-S.), & Le, T. P. N. (2019). Powerful Supply Chain: Apply Areas of Social and Science Management to Supply Chain Management to Enhance the Coordination of Supply Chain Participants
[19] Gaber, I. M. Hourly Electricity Price Prediction Applying Deep Learning for Electricity Market Management. Journal/Conference Proceedings Title, Volume (Issue),
[20] Sutaria, R., & Jain, R. (2020). Auto-Price Forecast: An Analysis of Car Value Trends.