Utilizing Machine Learning to Predict Agricultural Prices

Mrs. K. Vasumathi[1*], Dr. S. Selvakani[2], R. Hema Kumar[3]

[1]Assistant Professor, PG Department of Computer Science, Government Arts and Science College, Arakkonam – 631051.*

[2]Assistant Professor & Head, PG Department of Computer Science, Government Arts and Science College, Arakkonam – 631051.*

[3]PG Scholar, PG Department of Computer Science, Government Arts and Science College, Arakkonam – 631051.

Abstract

Agriculture forms the backbone of our country's economy, with farmers playing a vital role. However, in recent times, fluctuations in climatic conditions and other factors affecting prices have led to significant losses for farmers. Price prediction has emerged as a critical challenge in agriculture that can only be addressed by leveraging available data. By providing farmers with advance predictions, this application can help them increase profits and avoid substantial losses, thereby contributing to the growth of the national economy. To this end, we propose applying the ANN, Gradient Boosting, and Random Forest algorithms.

Keywords: Prediction, Machine Learning, Data Set, Agriculture Price prediction, ANN algorithm, Gradient boosting algorithm, and Random forest algorithm.

---

1. Introduction

Although machine learning algorithms have been applied successfully in various industries to solve prediction problems, there have been limited studies on their application to predict prices for agricultural products, especially vegetables. The agricultural sector is still lacking in technological advancements, and computer science tools such as big data and machine learning could play a crucial role in improving the situation. Most of the previous research on the application of machine learning in agriculture has focused on predicting weather, fertilizer quantities, and rainfall. There is a significant lack of epistemological research in the area of machine learning. As more techniques and technologies for data analysis become available, this area of research will become more pressing.

With the growing importance and utilization of machine learning, it is possible that philosophers of science will show an increasing interest in this topic in the near future. The goal of this study is to examine the existing research on predicting prices for agricultural products using machine learning algorithms.

The objective is to gain insight into the progression of the use of these algorithms, identify the most commonly utilized ones for agricultural price prediction, and analyze the research paradigms and performance metrics employed. In particular, Artificial Neural Network (ANN) has proven effective in learning fluctuations and providing more accurate forecasts.

Notably, ANN outperformed traditional statistical methods, such as Exponential Smoothing and ARIMA model, in predicting the rice exports of Thailand.

As the significance and utilization of machine learning continues to increase, it is possible that philosophers of science will become increasingly interested in this topic in the near future. The aim of this study is to investigate the current research on utilizing machine learning algorithms to predict prices for agricultural products. The objectives of the study are to gain a better understanding of the evolution in the use of these algorithms, identify the most frequently used algorithms for agricultural price prediction, and evaluate the research paradigms and performance metrics employed. Notably, Artificial Neural Network (ANN) has demonstrated effectiveness in learning fluctuations and generating more accurate forecasts. In fact, ANN outperformed traditional statistical methods, such as Exponential Smoothing and ARIMA model, in predicting the rice exports of Thailand.

2. Review of Literature

According to **DABIN ZHANG [1]** fluctuations in agricultural commodity prices have a significant impact on people's daily lives and the inputs and outputs of agricultural production. Accurate price forecasting is therefore crucial for agricultural authorities to make informed decisions. To depict agricultural commodity prices, 29 features are utilized, and three intelligent models are considered as potential forecast models: artificial neural network (ANN), support vector regression (SVR), and extreme learning machine (ELM). DABIN ZHANG [1] highlights the advantages of using complexity features to quantify chaos and measure the long-range dependence in a time series. Linearity features are also important in selecting appropriate models. Stationary features are utilized to assess the stationary of a time series, while periodicity features provide indications of the periodicity and seasonality of time series.

Kiran M. Sabu and T. K. Manoj Kumar [2] state that price fluctuations in agricultural commodities can have a negative impact on a country's GDP and can cause emotional and financial distress to farmers. Predicting prices may help the agriculture supply chain make informed decisions to minimize and manage the risk of price fluctuations Predictive analytics is expected to address the problems caused by unstable climatic conditions, global warming, and reduced agricultural production. Arcanum is a significant crop in India, with Kerala ranking second in production.

Other points mentioned by Kiran M. Sabu and **T. K. Manoj Kumar [2]** include the following:

- Classical time-series models, such as Holt-Winter's Seasonal Method and SARIMA models, and a machine learning model, LSTM, were compared to find the most parsimonious model.
- The ARIMA model is a Box Jenkins model used to predict or analyze stationary or non-stationary time-series data.
- The deep learning model LSTM was used for forecasting. Long-term temporal dependence is an essential aspect to consider in time-series data.
- The vanishing gradient problem is the primary limitation of traditional RNN, where the gradient diminishes and approaches zero.

Konstantinos G. Liakos [3] notes that since the publication of Kuhn's The Structure of Scientific Revolutions in 1962, the concept of a paradigm or worldview as a comprehensive framework that organizes our entire approach to existing in the world has become commonplace. This paper critically examines the positivist and non-positivist research paradigms in social science research and reveals that these two paradigms are in opposition to each other. According to **Konstantinos G. Liakos [3],** there are some advantages to the idea of a paradigm or worldview as an overarching framework. Despite the belief that a paradigm is beyond description and the understanding of the human intellect, it is believed that the intellect is more general than any world perception on which it takes its existing cognitive carriage. Therefore, it is necessary to expand individual awareness to express any essential way that an individual frames their world, as differences in epistemology, methodology, and supporting perspectives are often based on model assumptions.

Sue **L.T. McGregor [4]** emphasizes the importance of intellectual integrity, trustworthiness, and diversity in consumer scholarship, which is contingent upon researchers accounting for the philosophical underpinnings of their work. Many scholars tend to conflate methodology and method, which has prompted a discussion that distinguishes between these two concepts, identifies the four axioms of methodologies, describes two overarching research paradigms (positivism and post-positivism), and compares quantitative/qualitative with positivistic/post-positivistic approaches. Understanding these paradigms and methodologies is crucial in positioning consumer research and promoting its rigor and diversity.

Lovish Madaan and Ankur Sharma [5] state that price fluctuations in food can lead to difficulties for both consumers and producers, particularly in developing economies where perfect competition may not exist due to contextual factors. The authors examine the trading of onions and potatoes in India and propose a price forecasting model and an anomaly detection and classification system to identify instances of stock hoarding by traders.
Further, Lovish Madaan and Ankur Sharma [5] represented some advantages:

- High prices can increase the expenses of retail consumers, while low prices can reduce the incomes of farmer producers.

- In India, rainfall plays a significant role in price variation since most agricultural production relies on rain-fed irrigation.
- Commodity exchanges have been implemented in India to provide more predictable prices for farmers and buyers, but their reach and reliability remain uncertain.
- The mind system, which channels 60-80% of agricultural produce, requires regulatory mechanisms to be strengthened and information asymmetries to be addressed for these markets to function effectively.

Girish K. Jha and Kanchan Sinha [6] discuss the importance of food price forecasting for farmers, policymakers, and agribusiness industries in agriculture-dominated developing countries like India. Due to sparse and time-lagged data availability in these countries, time series forecasting models are commonly used. However, the recent advancement in Artificial Neural Network (ANN) modeling methodology provides a potential solution for price forecasting, given the availability of data. Additionally,

- Accurate price forecasting is crucial for commodity trading and price analysis.
- Forecasting models must have quantitative accuracy with small errors and turning point forecasting power.
- Agricultural commodity production and prices are often unpredictable due to natural calamities, such as droughts, floods, and pests and diseases, which pose a considerable risk and uncertainty in the process of price modeling and forecasting.

As per **Changshou Luo, Qingfeng Wei, Liying Zhou, Junfeng Zhang, Sufen Sun [7],** represented some advantages as:

- Predicting vegetable market prices is important as it is a significant source of income for farmers, and prices can be unstable and fluctuate rapidly.
- Unstable prices can hinder the sustained and steady development of vegetable production.
- Scientific methods, such as non-linear prediction methods like neural networks and genetic algorithms, can help to identify the change in vegetable prices.
- The prediction of market prices of agricultural products is complex and affected by multiple factors such as climate, supply, and demand.

As **per Konstantinos G. Liakos [8],** agriculture has witnessed the emergence of new scientific fields such as agric-technology, precision farming, and digital agriculture, which use data-intensive approaches to enhance agricultural productivity while minimizing environmental impact. Machine learning (ML) has evolved with big data technologies and high-performance computing, providing new opportunities to comprehend, quantifies, and predict data-intensive processes in agricultural production systems. The performance of ML models is evaluated using various statistical and mathematical models and is measured using performance metrics that improve over time with experience. The review highlights the use of artificial neural networks (ANNs) and decision trees (DT) as popular ML algorithms for classification and regression tasks.

3. Proposed Methodology

The proposed methodology for prediction involves the use of three machine learning algorithms: Linear Regression, Random Forest Regression, and Decision Tree Regression. The approach will follow the following steps:
Step 1: Initialize the dataset containing the training data for the wholesale price index.

Step 2: Select all the rows and column 1 from the dataset to "x," which is the independent variable.
Step 3: Select all of the rows and column 2 from the dataset to "y," which is the dependent variable.
Step 4: Fit the Decision Tree Regression (DTR), Support Vector Regression (SVR), or Linear Regression (LR) model to the dataset.
Step 5: Predict the new value using the trained model.
Step 6: Visualize the result and check the accuracy of the model.

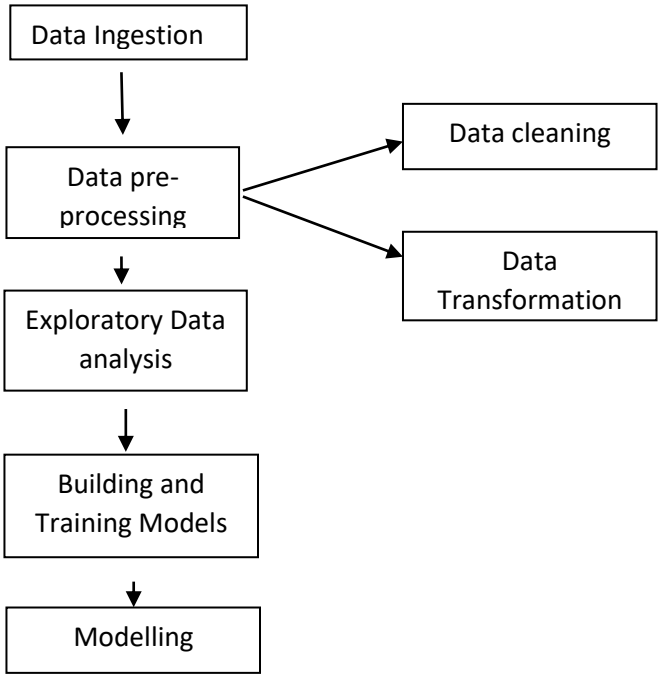

Figure 1. System Architecture

2.1 Advantages:

- The application provides farmers with a prediction in advance, helping them to increase their profits and avoid massive losses, ultimately contributing to the country's economy.
- It is also a time-efficient tool and improves the accuracy of the predictions.

2.2 Disadvantages:

- Four different crops were utilized for prediction.
- The time complexity was high.
- The prediction accuracy was relatively low.

2.3 Modules:

- Data Collection
- Data Pre-processing
- Feature Extraction
- Evaluation Model

Data Collection: The data utilized in this paper consists of a collection of product reviews obtained from records. The initial step involves selecting a subset of the available data to work with. Machine learning problems begin with data, ideally a large amount of data (examples or observations) for which the target answer is already known. Data for which the target answer is already known is referred to as labeled data.

Data pre-processing: Prepare your selected data by formatting, cleaning, and sampling it. • There are three common pre-processing steps for data: - Formatting: The data you have selected may not be in a format that is suitable for your work. For instance, it may be in a relational database and you need it in a flat file, or it may be in a proprietary file format and you need it in a relational database or a text file. - Cleaning: Cleaning data involves removing or fixing missing data. Some instances may be incomplete and may not carry the data you require to address the problem. As a result, they may need to be removed. Furthermore, some of the attributes may contain sensitive information, and they may need to be anonym zed or deleted from the data altogether. - Sampling: There may be far more data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

Feature Extraction: The subsequent step is Feature extraction, which involves reducing the number of attributes. While feature selection ranks existing attributes according to their predictive significance, feature extraction transforms the attributes themselves. The transformed attributes, or features, are linear combinations of the original attributes. After this, our models are trained using a Classifier algorithm. We utilize the classify module from the Natural Language Toolkit library in Python, using the labelled dataset we collected. The remaining labelled data will be used to evaluate the models. For the classification of pre-processed data, various machine learning algorithms were employed, including the popular Random Forest algorithm, which is commonly used in text classification tasks.

Evaluation Model: Model evaluation is a crucial aspect of the model development process. It aids in identifying the optimal model that represents our data and predicts how well the selected model will function in the future. Assessing model performance using the training data is not acceptable in data science since it can result in excessively optimistic and over fitted models. Two techniques for assessing models in data science are Hold-Out and Cross-Validation. Both methods employ a test set (not seen by the model) to assess model performance and prevent over fitting. The performance of each classification model is calculated based on its average and presented in a visual format, typically as a graph. Accuracy refers to the percentage of correct predictions for the test data and can be determined by dividing the number of correct predictions by the total number of predictions.

4. Experiments and Results

4.1 Testing

The purpose of software testing is to investigate and provide stakeholders with information about the quality of the product or service being tested. It also provides an objective, independent view of the software, enabling businesses to understand and appreciate the risks associated with implementing the software. Test techniques include the execution of a program or application with the goal of identifying software bugs. Software testing can be defined as the process of validating and verifying that a software

program/application/product meets the business and technical requirements that guided its design and development, and that it functions as expected and can be implemented with the same characteristics.

4.2 Testing Methods

4.2.1 Functional Testing

Functional testing systematically demonstrates that the functions being tested are available as specified by the business and technical requirements, system documentation, and user manuals. The focus of functional testing is on the following items:

- Functions: The identified functions must be exercised.
- Output: The identified classes of software outputs must be exercised.
- Systems/Procedures: The system should work properly.

4.3 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

Figure 2. Actual vs Fitted Values for Price

Result

Data mining is a process to extract knowledge from existing data. It is used as a tool in banking and finance, in general, to discover useful information from the operational and historical data to enable better decision-making. It is an interdisciplinary field, the confluence of Statistics, Database technology, Information science, Machine learning, and Visualization. It involves steps that include data selection, data integration, data transformation, data mining, pattern evaluation, knowledge presentation.

Conclusion

The research aims at predicting the price and forecast through backend application and it is runs on efficient machine learning algorithms and technologies having a good accuracy. The training datasets so obtained provide the enough insights for predicting the

appropriate price and demand in the markets. Thus, the system helps the farmers in reducing their difficulties and stops them by attempting suicides.

References:

1. Aliyu, A., Bello, M., Kasim, R., Martin, D.[3] (2014). Positivist and non-positivist paradigms in social science research: Conflicting paradigms or perfect partners? Journal of Management and Sustainability, Vol. 4, No.3, pp. 79–95.
2. Jha, G., Sinha, K.[6] (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. Agricultural Economics Research Review, Vol. 26, No. 2, pp. 229-239.
3. Jha, G., Sinha, K.[7] (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. Agricultural Economics Research Review, Vol. 26, No. 2, pp. 229-239.
4. Jha, G., Sinha, K.[8] (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. Agricultural Economics Research Review, Vol. 26, No. 2, pp. 229-239.
5. McGregor, S. Murnane, J.[4] (2010). Paradigm, methodology and method: Intellectual integrity in consumer scholarship. International Journal of Consumer Studies, Vol. 34, No. 4, pp. 419- 427.
6. Sabu, K. M., Kumar, T. K. M. [2] (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. Procedia Computer Science, Vol. 171, pp. 699-708.
7. Xiong, T., Li, C., Bao, Y., Hu, Z., Zhang, L.[5] (2015). A combination method for interval forecasting of agricultural commodity futures prices. Knowledge-Based Systems, Vol. 77, pp. 92–102.
8. Yashavanth, B., Singh, K., Paul, A., Paul, R.[9] (2017). Forecasting prices of coffee seeds using vector autoregressive time series model. Indian Journal of Agricultural Sciences, Vol. 87, No. 6, pp.754–758.

Mrs.K VASUMATHI Working as an Assistant Professor Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed S.E.T (Computer Science), M.Phil (Computer Science), M.Sc (university VI Rank) at Thiruvalluvar University, Vellore, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 13 years in Teaching. She is having hands on experience in Research contributions and Student Counseling. She has published more than 18 papers in the peer reviewed International Journals. Besides presented more than 35 papers in the National and International Conferences. She has published 2 books. Her interest areas are Operating Sysyem, Big Data, Machine Learning**.**

Dr.S.Selvakani Working as an Assistant Professor and Head in the Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed Ph.D (Computer Science), M.Tech (Computer Science and Engg), M.Phil (Computer Science), MCA (university III Rank) at Manonmanium Sundaranar University, Tirunelveli, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 20 years in Teaching, Administration/ Training & Development. She is having hands on experience in general administrative activities, Research contributions and Student Counseling. Cambridge University has recognized and awarded certification of "Teachers and Trainers" for remarkable achievement in the field of Teaching. She has published more than 48 papers in the peer reviewed International Journals and more than 6 papers in National Journals. Besides presented more than 90 papers in the National and International Conferences and guiding several Ph.D Scholars in Computer Science presented more than 112 papers in National and International Journals, Conference and Symposiums. She has published 4 books and 6 Patents. Her interest areas are Data Science, Big Data, Machine Learning and Network Security.

R Hema Kumar has received his Bachelor's Degree from Thiruvalluvar University College of Arts and Science and doing his Masters in Government Arts and Science College, Arakkonam. His research interests are Data Science, Big Data and Network Security.