

Onion Price Prediction Based on Artificial Intelligence

R. Nalini ^{1,*}, K. Sountharya ¹, R. Vishnu Priya ¹, R.V. Punitha ¹

¹ Department of Computer Science and Engineering, K. Ramakrishnan College of Engineering, Tiruchirappalli, Tamil Nadu, India.

*Corresponding author E-Mail ID: nalinicsere@gmail.com

DOI: <https://doi.org/10.34256/irjmt2043>

Abstract

India being a horticulture nation, its economy prevalently relies upon horticulture yield development and agroindustry items. Information Mining is a developing examination field in onion yield investigation. Yield forecast is a significant issue in horticultural. Any rancher is keen on knowing how a lot yield he is going to anticipate. Break down the different related properties like area, pH esteem from which alkalinity of the dirt is decided. Alongside it, level of supplements like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is utilized along with the utilization of outsider applications like APIs for climate and temperature, sort of soil, supplement estimation of the dirt in that locale, measure of precipitation in the district, soil organization can be decided. Every one of these traits of information will be examined, train the information with different appropriate AI calculations for making a model. The framework accompanies a model to be exact what's more, exact in foreseeing onion yield and convey the end client with appropriate proposals about required manure proportion in light of barometrical and soil parameters of the land which improve to build the harvest yield and increment rancher income.

Keywords: Yield, AI, APIs, Horticulture yield, Soil parameters

1. Introduction

Farming makers are progressively inclined to climate hazard and value dangers than some other segment and explicitly little and peripheral ranchers who regularly does the trouble offer of their produce. Since a decade ago, high unpredictability in costs of agrarian products has been a worry for the two makers and customers. Ranchers are in trouble under both the circumstances of harvest disappointments and guard preparations. During guard creation, due to excess in the market, ware costs tumble to the absolute minimum hitting the ranchers' salary hard. Significant expense inconstancy in agribusiness concurs significance to solid value anticipating systems for ranch produce. The value figures are useful for ranches, policymakers what's more, agri-business enterprises. The fundamental reason for farming ware cost determining is to permit makers to settle on better-educated choices and to oversee value chance. Utilizing time arrangement information in horticulture, nonstop endeavors are made by the analysts to foresee the costs utilizing numerous direct and nonlinear gauging models.

As of now, Artificial insight/Machine learning models have set up themselves as genuine contenders to traditional measurable models in the estimating network. For value expectations, explicitly AI gives a one-of-a-kind method for joining specialized examination and key investigation strategies. While specialized examination exclusively takes a gander at authentic costs, central examination comprises of understanding the outer what's more, inner elements that

impact the cost of a specific stock. Independently they have qualities, specialized investigation gives precise transient forecasts while basic examination is progressively fit toward long haul standpoint. By consolidating the two, more exactness in expectations can be accomplished. Subsequently a pilot venture has been taken ready for anticipating the costs of and Onion (Rabi) applying AI procedures in this way utilizing on verifiable cost and information factors.

2. Related Work

The existing system, build up a basic and reasonable anticipating model at month to month residential onion costs in view of SVR (Support Vector Regression). For the reasons for this examination, they used stock information gave by the ERP arrangement of APC. APC is a mid-dispersion specialist in the horticultural division. The administration of South Korea (henceforth, Korea) has put forth attempts in balancing out horticultural item costs (Jeong et al., 2017). Agrarian item costs display high changeability, particularly the costs of vegetables, for example, onion, cabbage, what's more, radish (Jeong et al., 2017). It is notable that exceptionally fluctuating costs negatively affect both ranchers and consumers. Despite the way that agrarian ware costs are of extraordinary criticalness in the farming division, it is hard to acquire precise value expectations. Data gave by mid-dissemination specialists is fundamental for accurate horticultural product value guaging (Jeong et al., 2017). This is ascribed to the way that agrarian products are disseminated to the discount and the retail advertise through mid-appropriation specialists for example, APC. By getting data from mid-circulation operators, for example stock sum and shipment sum, they anticipate costs all the more precisely. Therefore, they center around APC ERP information, particularly on shipment sum. In spite of the fact that getting shipment information from APC ERP isn't clear, in the event that we could give proof that the information is vital for estimating agrarian item value, government would give more consideration to dispersion of APC in horticultural field.

This model has the capability of being valuable and important to the rural area and government chiefs at estimating costs and distinguishing the effect of market basics on horticultural item costs. The support vector machine (SVM) combines concepts from abstract Hilbert spaces with modern optimization techniques (Cui & Curry, 2005). SVM is well known as an effective approach for solving classification problems (Heo, 2013). Furthermore, SVMs can effectively handle regression and forecasting tasks. SVMs fall into two categories, SVC (Support Vector Classification) and SVR. When dealing with regression problems, SVR should be used (Alex & Bernhard, 1998). In Existing system, Support Vector Machine (SVM) Algorithm is used. SVM algorithm is supervised learning algorithm. This algorithm is used for classification and regression. It is also referred as Kernel trick which is to transform your data and based on these transformations it finds an optimal boundary between possible outputs. But the disadvantage of SVM is not suitable for large dataset. It also not performs well when target classes are overlapping.

Data mining is another way toward extricating significant and valuable data from huge arrangements of information [1]. Data mining in farming is a novel investigate field. Ranchers are not just reaping vegetables and crops yet in addition collecting huge measure of information. Information mining gives the procedure to change this information into valuable data for basic leadership. Vegetable value changes quick furthermore, temperamental which has incredible effect in our everyday life. Vegetable cost has qualities, for example, high nonlinear and high clamor. In this way, it is difficult to anticipate the vegetable cost. Information mining order procedures can be utilized to build up an inventive model to anticipate the market cost of particular ware. Value expectation is exceptionally helpful in horticulture for anticipating the market cost for the particular items and furthermore helpful for ranchers to design their yield development exercises so that they could bring more cost in the market. Government can utilize the market conjecture cost for arranging and execution of farming advancement projects to balance out the market cost for the particular ware. Government can likewise take choice whether to permit or not to fare and import

of particular items. Buyers can utilize this value forecast for their everyday way of life arranging. This imaginative application isn't helpful for ranchers and shoppers yet in addition helpful for farming arranging, surrounding polices and plans in horticulture and market arranging. Information mining arrangement method, for example, Neural Network plays a significant job in non-straight time arrangement expectation [2-5]. There are numerous sorts of expectation strategy on premise of Neural System, among them the utilization of BP Neural Network calculation is most significant one.

3. Proposed System

A general methodology for AI ventures has been received for this undertaking. The following stages were followed in figure 1.

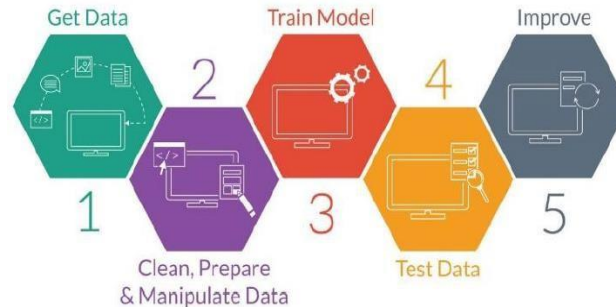


Figure 1. Overview of proposed system

Identify determinants and acquire data: Interviews with domain experts, identification of influencing factors and delivery of historical data from key stakeholders.

- i) Prepare data: Clean and transform the downloaded data into a machine-readable format.
- ii) Train model: Developing multiple price prediction algorithms.
- iii) Test model: Testing the output of the algorithms and shortlisting best performing algorithms.
- iv) Deploy and live testing: Deploy the algorithm for live testing over a period and further improvement using key learnings.

A) Training Algorithm

The presentation of various ML calculations firmly relies upon the size and structure of the info information. In this way, the right decision of calculation regularly stays misty except if we try out our calculations straightforwardly through plain old experimentation for onion, the all-out dataset comprised of data points from 2012 onwards.

Our preliminaries comprised of various blends of the accompanying three parameters:

1. Class of calculations: Time arrangement guaging, choice trees and progressed relapse calculations
2. Number of value determinants included while preparing: Ranging from utilizing 1 to 14 determinants
3. Information from number of mandis for preparing: Training the calculation on information running from 1 to 30 mandis.

Backpropagation algorithm is supervised learning for Multi-Layer feed forward Network in Artificial Intelligence. Supervised Models are getting trained in Labelled Dataset (i.e.). Both Input and Output. The main Objective of Backpropagation algorithm is to develop learning algorithm for Multilayer feed network. It is used to find local minimum error of the function. This algorithm

works by computing gradient loss function with respect to weights by chain rule. Backpropagation starts with Random weights and goal is to adjust the error until artificial neural network learns the training data.

The decision for the classes of calculations depended on a writing audit for cost determining techniques. For time arrangement determining, we chose ARIMA (Auto Regressive Coordinated Moving Average), for choice trees we applied Random woodland and LASSO (least outright shrinkage and determination administrator), SVM (bolster vector machine) and GLM (summed up direct model) for Regression. The clarification of the different calculations is given in informative supplement A. The Root Mean Square Error (RMSE) was determined during the preparation and testing of every calculation. RMSE is the standard deviation of the residuals (forecast mistakes). Residuals are a proportion of how a long way from the relapse line information focuses are; In other words, it discloses to us how focused the information is around the line of best fit.

$$RMSE = \sqrt{(f - o)^2}$$

Where f = anticipated costs and o = watched costs or actuals.

B) Testing Algorithm

Model evaluation is used to test the trained data. It estimates the generalization accuracy of the model on future data. To evaluate the performance their methods are categorizes into two: Holdout and Cross-validation. Holdout is used to test a model on different data than it was trained. This provide unbiased learning of learning performance. Cross-validation involves partitioning the original observation dataset into training data's, used to train data and an independent set used to evaluate the analysis. Most common Cross-validation is k-fold cross-validation. It takes k equal sizes subsamples called folds. Here sequence of models is trained. The first model is trained using the first fold as the test set, and remaining are used as training set. This is repeated for each of these k splits of data and the estimation of accuracy is averaged overall to get effectiveness of our model.

For Onion, the preparation set included data points from January 2012 to January 2017, the testing was accomplished for the period from February 2018 to January 2019. For Onion, the preparation set involved data points from January 2012 to January 2017, the testing was accomplished for the period from February 2018 to January 2019.

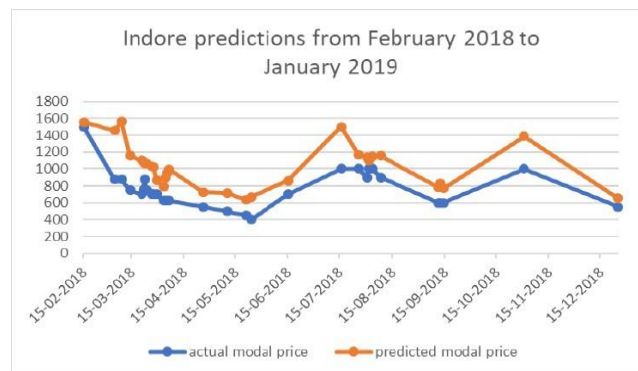


Figure 2. Onion predictions for Indore from February' 18 till January' 19

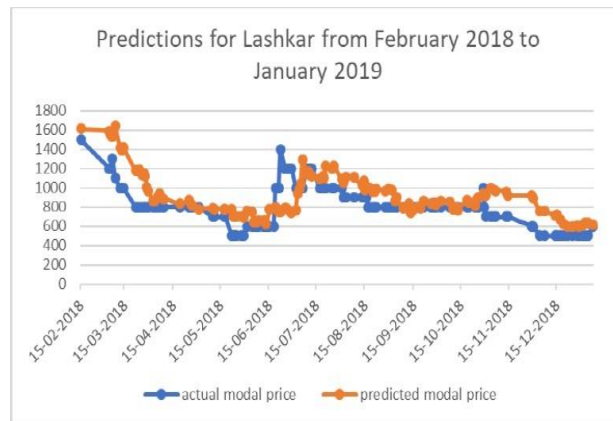


Figure 3. Predictions for Lashkar from February’18 to December’18

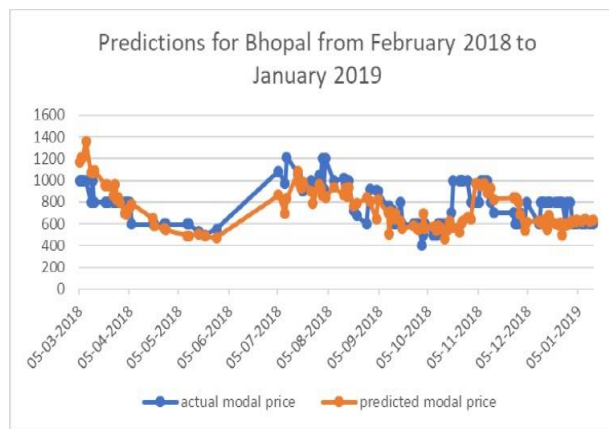


Figure 4. Predictions for Bhopal from February’18 to January’19

C) Overview of The Various Algorithms (Onion) Tested During the Development Phase for Onion.

The below table 1. shows the overview of various algorithms which is tested during the development phase of onion.

Table 1. Overview of various algorithms

Class of algorithm	Periodicity of input dataset	Model type	Number of price determinants	Number of mandis	RMSE (Root mean squared error)	Additional remarks
Time Series Forecasting	Daily	ARIMA (2-weekly)	Univariate (only price)	1 Mandi	295.28 Indore	Prediction provided two weeks ahead
	Daily	ARIMA (monthly)	Univariate (only price)	1 Mandi	498.93 Indore	Prediction provided a month ahead
Advanced regression	Daily	Lasso (2-weekly)	12 variables	15 Mandis	219.72	Prediction provided 2 weeks ahead
	Daily	Lasso (monthly)	12 variables	15 Mandis	316.57	Prediction provided one month ahead

4. Experimental Results

Table 2. Onion predictions (monthly) for 14th November'2018

Name of Mandi	Actual Modal Price	Predicted Modal Price	Error %
Bhopal(F&V)	700	825.30	15.18%
Dewas(F&V)	600	845.66	29.04%
Haatpipliya	1200	1041.22	15.24%
Harda(F&V)	850	1146.21	25.84%
Khandwa(F&V)	600	624.77	3.96%
Lashkar(F&V)	700	918.58	23.79%
Sagar(F&V)	700	1276.14	45.14%
Sanwer	450	694.45	35.20%
Shajapur(F&V)	800	725.66	10.24%

Table 3. Onion predictions (monthly) for 15th December 2018

Name of Mandi	Actual Modal Price	Predicted Modal Price	Error %
Bhopal(F&V)	800	639.29	20.08%
Deori	600	549.44	8.42%
Dewas(F&V)	300	605.20	101.73%
Khandwa(F&V)	550	409	25.63%
Lashkar(F&V)	500	726.33	45.26%
Neemuch	450	513.71	14.15%
Sanwer	450	543.26	20.72%

- a) AI for value expectations: The inward outcomes alongside the live testing results have demonstrated that AI can be applied to give cost expectations. Our methodology is exceptional as far as its capacity to give forecasts at a mandi level. The value forecasts for Soybean were given until end of January'19. This gave us an example of forecasts for a quarter of a year. Inside this period, we could find patterns with respect to blunder rates (for ex. on the off chance that certain mandis which show reliably higher mistake rates, certain months show higher rates).
- b) Recorded information assortment: The way toward downloading individual datasets going from 2008 onwards is tedious. The greater part of the information is downloaded from agmarknet.gov.in, which sadly has constrained APIs for mechanized downloading. That said once the underlying dataset is made, it requires least exertion to refresh the authentic database.
- c) Accessibility of preparing information: We have seen that the agmarknet.gov.in endures from ordinary personal time and information concerning costs isn't reliably refreshed over all Mandis. This exceptionally constrains our capacity to anticipate costs for all Mandis for shorter periods (15-day interims).

- d) Long haul scattering situation: We are right now investigating approaches to robotize the whole procedure, directly from chronicled information download, refreshing the dataset at a week after week premise to giving the forecasts as an API.
- e) Reasonableness of yields for value forecasts: The pilot additionally approved that accessibility of dependable information significantly affects the precision of the calculation. For soybean there was abundant information while onion information was rare. We suspect the equivalent for most plant crops

A. Monthly Simulation and Prediction Analysis of BPNN

Taking former four-month data of tomato as input and later one-month data as output the code is developed using MATLAB. BPNN is constructed using Jan 2009 to May 2011 monthly price data and later month’s data are used to test the model. The number of hidden neurons is 5. Optimization target is 0.001.

Table 4. Monthly predicted price with error

Month	Actual	Predicted	Error %
1	14	15.0036	7.168571
2	13	13.2024	1.556923
3	13	12.458	4.169231
4	9	9.1115	1.238889
5	13	12.4633	4.128462
6	18	17.4878	2.845556
7	26	25.5212	1.841538

For figuring the supreme blunder between estimated esteem what's more, anticipated worth. We utilize the mistake rate recipe, $(\text{Genuine value} - \text{Predicted cost}) / \text{Actual value} * 100$ (2) Right now saw that the outright mistake of month to month value forecast is inside 10%, So the exactness is up to 90%.

B. Week After Week Prediction Analysis of BPNN

Taking multi week's information of tomato as info and later one-week information as yield, we have composed program in MATLAB. BPNN is developed utilizing past 135 weeks from Jan 2009 value information and later week's information are utilized to test the model. The number of shrouded neurons is 4. Improvement target is 0.001.

The outcomes are as per the following,

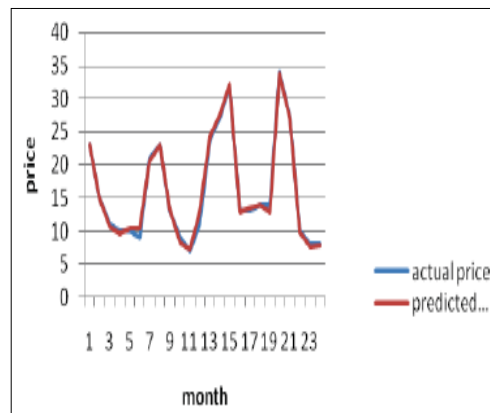


Fig 5. Monthly price prediction

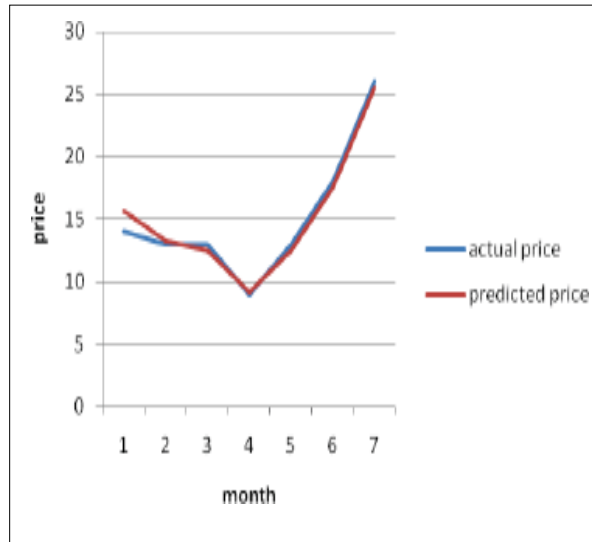


Fig 6. Prediction result

Table 5. Weekly predicted price with error

Week	Actual price	Predicted price	Error %
1	26	20	23.07692
2	31	26	16.12903
3	34	28	17.64706
4	25	24	4
5	18	20	11.11111
6	18	18	0
7	14	17	21.42857
8	11	13	18.18182
9	9	10	11.11111
10	10	10	0

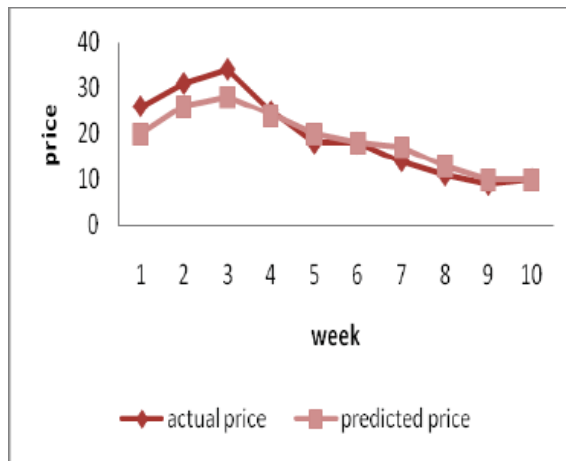


Fig 6. Weekly Prediction result

5. Conclusion

In this paper, we have exhibited that AI procedures can be effectively applied to gauge costs if the dependable noteworthy information at the different cost determinants are accessible. The models were prepared utilizing 10 years of authentic information (for all impacting factors) from 2008 to 2018. For soybean, a normal precision of 95% has been accomplished and exhibited during the "live" testing for the forecasts (for 15 days interim) across Mandis beginning from mid-October'18. The relating mistake rate extended from 8.35% to 0.01%. For onion in any case, the outcomes were less reassuring. A normal exactness of 76% was accomplished. We contribute this to extraordinary unpredictability and an absence of preparing information for onion. The preparation time frame incorporated the choice of right sort of calculation from various classes. Rope (a class of relapse strategies) has demonstrated to be the most reasonable calculation for the 15-and 30-day interims.

References

- [1] J. Abello, P. M. Pardalos, M. Resende, (2002) Hand Book of Massive Datasets, *Massive Computing*, Springer, United States, 1223.
- [2] Chapgshou Luo, Qingfeng Wei, Liying Zhou, Jungeng Zhang, R. Suien Sun, (2011) Prediction of Vegetable Price Based on Neural Network and Genetic, Algorithm, *In International Conference on Computer and Computing Technologies in Agriculture*, Springer, Berlin, 672-681
- [3] Zahir Haiderkan, Tasnim Sharmin Alin, Akter Hussain, Price Prediction of Share Market Using Artificial Neural Network (ANN), *International Journal of Computer Applications* 22 (2011) 42-47.
- [4] K. G. Akintola, B.K. Alese, A. F. Thompson, Timeseries Forecasting with Neural Network –A Case Study of Stock Price of Intercontinental Bank Nigeria, *International Journal of Recent Research and Applied Studies*, 9 (2011) 467-472.
- [5] T. Jayalskshmi, A. Santhakumar, Statistical Normalization and Back Propagation for Classification, *International Journal of computing Theory and Engineering*, 3 (2011) 1793-8201.
- [6] J. P. Donate, X. Li, G. G. Sánchez, A. S, de Miguel, Time Series Forecasting by Evolving Artificial Neural Networks with Genetic Algorithms, Differential Evaluation and Estimation of Distribution Algorithms, *Neural Computing & Applications*, 22 (2013) 11-20.
- [7] M. Shukla, S. Jharjharia, Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation, *International Journal of Information Systems and Supply Chain Management (IJISSCM)* 6 (2011) 105-119.
- [8] T. G. Manohar, V.V. Reddy, Load Forecasting by A Novel Technique Using ANN, *ARPV Journal of Engineering and Applied Sciences*, 3 (2008) 19-25.

Funding

This study was not funded by any grant

Conflict of interest

None of the authors have any conflicts of interest to declare.

About The License

The text of this article is licensed under a Creative Commons Attribution 4.0 International License

Cite this Article

R. Nalini, K. Sountharya, R. Vishnu Priya, R. V. Punitha, Onion Price Prediction Based on Artificial Intelligence, International Research Journal of Multidisciplinary Technovation, Vol 2, Iss 4 (2020) 11-20.

DOI: <https://doi.org/10.34256/irjmt2043>