

Onion Yield Prediction Based on Machine Learning

A.Selvi^a, D.Dharani^b, T.Gobika^c, S.Harini^d and N.Nithya^e

^a

Department of Computer Science and Engineering , M.Kumarasamy College of Engineering , Karur, TamilNadu, India-639113.

^{b,c,d,e} Department of Computer Science And Engineering, M.Kumarasamy College of Engineering , Karur, Tamilnadu, India-639113.

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Indian economy is mostly based on agriculture but the price of each vegetables and fruits varies in day today life. In recent days, the price of onion is increasing wisely, to predict the future price of onion by using past data and also the water level need for a particular crop using current dataset and also by using climate changes. Using Machine learning, collect the dataset of weather report and past and current price of vegetables (e.g. onion) and water level prediction of plants. By analyzing these dataset the price of vegetables and the water level required for particular plant on particular day according to the climate change are analyzed and predicted. By using machine learning algorithm like Support vector machine, stemming, and tokenization. This will be helpful for the cultivators to crop a plant according to the weather and it will be useful when the water level needed for the particular plant is known previously. And these will also help to bring customer and farmer will gain more. The water level will also help to enhance the harvest yield and also will produce good crops all the time. The cultivator will gain more and the harvesting will also be more when compared to other systems.

Keywords: machine learning, linear regression algorithm, vegetable price prediction, weather analysis and water level prediction.

1. Introduction

Vegetables from farmers are in advanced minded due to climatic changes are riskier than division and indirectly outline the farmers who routinely does the hardwork for the yield of products[7,8,9]. For the past decade ago there is a high demand in price prediction[3,5]. The livestock farmers are in issues under harvest of vegetables[1,3,4,10,11,12]. Due to this farmers are gaining low income[2]. Onion and vegetables are more important to lead a healthy life. If vegetable price steadily increases poor people can't buy it[8,9,10]. So, they can't lead a healthy life.

Price prediction is a useful idea for consumers as well as farmers[3,4,5]. Here, Machine learning helps businesses to set a pricing manner that develops loyalty among people[1,2,3]. Yielding forecast is a important issue in agriculture. Vegetable value changes according to weather conditions[9,14,15,20]. By using Machine learning algorithm we can predict the price of vegetables next day by comparing it with previous price rate of vegetables[27,28,29,30,31,32,33]. For this we used the following algorithms like Support Vector Machine, Stemming, and Tokenization.

2. Literature survey

[1] The proposed system is about the collection of massive dataset collection, the problems in collecting the dataset and also includes collecting the dataset both in the internet as well as in the traditional way. The massive development in computer technology brings need of most of the data so they produced the system of collecting more number of dataset.

[2] The project predicted the market price of vegetables with the help of back propagation neural network and genetic algorithm network model. They have collected about 84 models and performed training and testing. The integrated model predicted and produced good results.

[3] They proposed the prediction of price only and have used back propagation neural network the vegetable taken was tomato and they have analyzed the features of tomato and predicted the price. The result was taken in the form of percentage.

[4] This paper predict the values of shares so that the buyers will be able to acquire their shares in the appropriate time. They have used error back propagation algorithm to train the neural network so that the price of stock price will be predicted shortly.

[5] They proposed the prediction of price and they have taken tomato as an example and have used back propagation neural network and they have analyzed the features of tomato and predicted the price to help the cultivators as well as the government. The result was taken in the form of percentage.

[8] By collecting the dataset of weather and climatic conditions the weather is prediction in this paper. They have used back propagation neural network for initialization and ANN based method.

[9] This paper describes about the evolution of bigdata it consist of economics of big data.

3. Proposed System

In proposed system, price prediction and water level management has been implemented using Machine Learning Algorithm figure 3.1. This algorithm has been implemented because of its accuracy. The preprocessing of data is done using Stemming and Tokenization. The linear regression algorithm is used to predict the price and water required for a particular plant for a specific period.

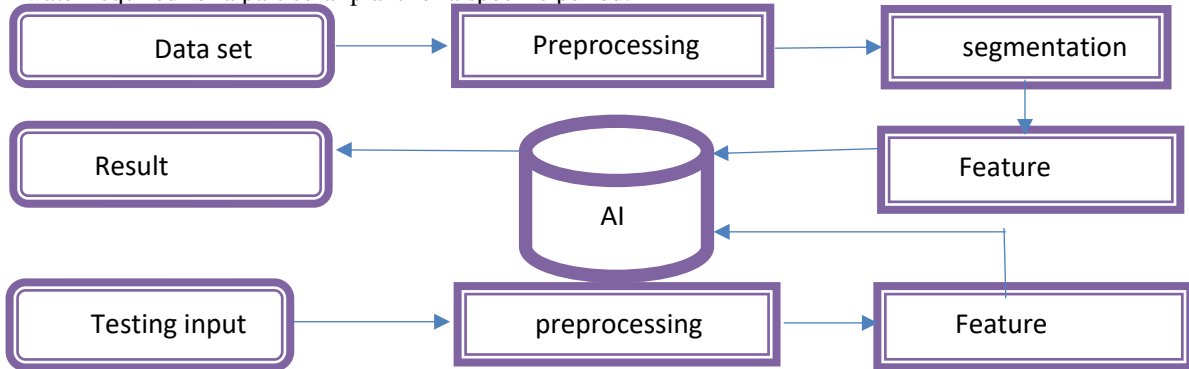


Figure 3.1: Workflow for training the machine

3.1.A) Dataset collection

Data collection is a process to collect the data in the form of structured and unstructured data to make a database. Here the dataset is collecting the data about weather, onion prize, and water level. The data stored as the database and the machine is trained after pre-processing.

3.2.B) Preprocessing of data

Data pre-processing is a way of converting data from the raw form to a much more usable or desired form. To make the data more significantly standardizing, rescaling,finalizing, one hot encoding, and label encoding. Here's the list of contents for this module.

3.3.C) Feature extraction

The Features with a high percentage of missing values shown in figure 3.2.Collinear (highly correlated) features.Features with zero importance in a tree-based model.Features with low importance.Features with a single unique value.

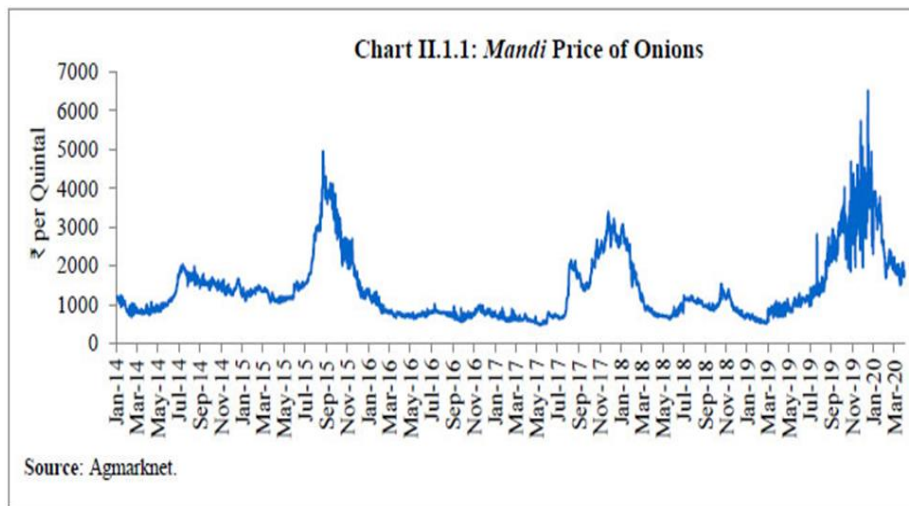


Figure 3.2: Price prediction of onion upto march 2020

3.4.D) Training the machine

Machine learning is a field of computer science that defines making the machine to learn from past historical data and after training and testing correct output is predicted in the form of pictorial representation or in the theoretical representation.By using the dataset that we have collected the machine is trained and tested for producing the output.

3.5.E) Classification

Firstly, Load Python packages and the data are pre-processed. After preprocessing the subset of the data is collected. Then splitting the data into training and testing dataset. After dividing the data into training and testing the data is tested and trained then build a random forest classifier. Then predict the price of onion then check the accuracy of the model. Finally, check the feature and importance's.

3.6F) Literature Results

The sample result shown in table 3.1

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%
HARDA(F&V)	850	1146.21	25.84%
KHANDWA(F&V)	600	624.77	3.96%
SAGAR(F&V)	700	1276.14	45.15%
SANWER	450	694.45	35.20%
SHAJAPUR(F&V)	800	725.66	10.24%

Table 3.1: Onion predictions (monthly) for 15th December 2018

3.7.G) Prediction analysis

The main execution of the algorithm to analyze and predict the prize and water level for the given dataset. In the existing projects they have predicted the price of vegetable in monthly basis but here the prices are predicted in the daily basis.

The outcomes are as per the following figure 3.3,



Figure 3.3 sample price prediction

4. Conclusion

By predicting and analysis of demand and water level management and weather prediction plays a vital role in cultivating the crops. If we predict the weather and water level it will help the cultivators to cultivate the crops. Then the major part of our project is price prediction will increase the yield if their yield increases means farmers will gain more than the present income this will also help the people to manage their expenses if the price was predicted earlier means they will buy the vegetables before and the consumption also increases. This project will give the accurate price so that there is no deviation in the thoughts of people about the price increase. This will definitely help the people to manage the vegetable expenses it can be implemented for all the vegetables as of now onion is one of the most demanded vegetables so we have particularly taken that if it is implemented for all the vegetable means it will be most useful to both the consumers and the farmers.

References

1. Abello. J. Pardalos PM, Resende M, Hand book of massive datasets. Vol. 3 of the Kluwer Series on Massive Computing, 2002Kluwer, New York.
2. Murugesan, M., Thilagamani, S., "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network", Journal of Microprocessors and Microsystems, Volume 79, Issue November 2020, <https://doi.org/10.1016/j.micpro.2020.103303>
3. ZahirHaiderkan, TasnimSharminAlin, & Md. Akter Hussain "Price prediction of share market using Artificial Neural Network". International Journal of Computer Applications Vol 22-No2-may2011
4. Akintola K.G., Alese B.K. & Thompson A.F. "Timeseries forecasting with neural network –a case study of stock price of intercontinental bank Nigeria" IJRRAS Dec2011.
5. T.Jayalskshmi, Dr.A.Santhakumar "Statistical Normalization and Backpropogation for classification" International Journal of computing Theory nad Engineering, Vol 3- No1-Feb2011.

6. Thilagamani, S., Nandhakumar, C. .” Implementing green revolution for organic plant forming using KNN-classification technique”, *International Journal of Advanced Science and Technology*, Volume 29 , Issue 7S, pp. 1707–1712
7. Manish Shukla, Sanjay Jharjhar “Applicability of ARIMA models in Wholesale vegetable market, An investigation” Preceeding of the 2011 “International Conference on Industrial Engineering and Operations Management” KualaLumour, Malasia, Jan22-24,2011.
8. Thilagamani, S., Shanti, N.,” Gaussian and gabor filter approach for object segmentation”, *Journal of Computing and Information Science in Engineering*, 2014, 14(2), 021006, <https://doi.org/10.1115/1.4026458>
9. Rhagini, A., Thilagamani, S. ,”Women defence system for detecting interpersonal crimes”,*International Journal of Advanced Science and Technology*, 2020, Volume 29,Issue7S, pp. 1669–1675
- A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.
10. L. A. Gabralla, R. Jammazi, and A. Abraham, "Oil price prediction using ensemble machine learning," in 2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE), 2013, pp. 674-679: IEEE. [11] V. Anandhi, R. M. J. I. j. o. f. c. Chezian, and communication, "Support vector regression to forecast the demand and supply of pulpwood," vol. 2, no. 3, p. 266, 2013.
11. K.Deepa, S.Thilagamani, “Segmentation Techniques for Overlapped Latent Fingerprint Matching”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN: 2278-3075, Volume-8 Issue-12, October 2019. DOI: 10.35940/ijitee.L2863.1081219.
12. D. Sinta, H. Wijayanto, and B. J. A. M. S. Sartono, "Ensemble knearest neighbors method to predict rice price in Indonesia," vol. 8, no. 160, pp. 7993-8005, 2014.
13. Kumararaja, V., Deepa, K.,” Pap smear image classification to predict urinary cancer using artificial neural networks” , *Annals of the Romanian Society for Cell Biology*, ISSN:1583-6258, Vol. 25, Issue 2, 2021, Pages. 1092 – 1098
14. M. Rafieisakhaei, B. Barazandeh, and M. Tarrahi, "Analysis of supply and demand dynamics to predict oil market trends: A case study of 2015 price data," in *SPE/IAEE Hydrocarbon Economics and Evaluation Symposium*, 2016: Society of Petroleum Engineers.
15. Santhi, P., Priyanka, T.,Smart India agricultural information reterival system, *International Journal of Advanced Science and Technology*, 2020, 29(7 Special Issue), pp. 1169–1175
16. J. M. Keller, M. R. Gray, J. A. J. I. t. o. s. Givens, man., and cybernetics, "A fuzzy k-nearest neighbor algorithm," no. 4, pp. 580-585, 1985.
17. J. J. I. J. o. A. E. R. COE, "Performance comparison of Naïve Bayes and J48 classification algorithms," vol. 7, no. 11, p. 2012, 2012.
18. Santhi, P., Mahalakshmi, G., Classification of magnetic resonance images using eight directions gray level co-occurrence matrix (8dglcm) based feature extraction, *International Journal of Engineering and Advanced Technology*, 2019, 8(4), pp. 839–846
- A. J. I. J. o. E. T. Pradhan and A. Engineering, "Support vector machine-A survey," vol. 2, no. 8, pp. 82-85, 2012.
19. M. KangaraniFarahani and S. Mehralian, "Comparison between artificial neural network and neuro-fuzzy for gold price prediction," in 2013 13th Iranian Conference on Fuzzy Systems (IFSC), 2013, pp. 1-5: IEEE.
20. Pandiaraja, P, Vijayakumar, P, Vijayakumar, V & Seshadhri, R 2017, ‘Computation Efficient Attribute Based Broadcast Group Key Management for Secure Document Access in Public Cloud’, *Journal of Information Science and Engineering*, 33, No. 3, pp. 695-712
21. T. Zhao, Y. Chen, A. Ray, and D. Doll, “Quantifying almond water stress using unmanned aerial vehicles (uavs): correlation of stem water potential and higher order moments of non-normalized canopy distribution,” in *ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2017, pp. V009T07A058–V009T07A058.
22. R. Nasi, E. Honkavaara, P. Lyytik ” ainen-Saarenmaa, M. Blomqvist, ” P. Litkey, T. Hakala, N. Viljanen, T. Kantola, T. Tanhuanpa” a, and ” M. Holopainen, “Using uav-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level,” *Remote Sensing*, vol. 7, no. 11, pp. 15 467–15 493, 2015.
23. T. Zhao, H. Niu, A. Anderson, Y. Chen, and J. Viers, “A detailed study on accuracy of uncooled thermal cameras by exploring the data collection workflow,” in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III*, vol. 10664. International Society for Optics and Photonics, 2018, p. 106640F.

24. Vijayakumar, P, Pandiaraja, P, Karuppiah, M & Deborah, LJ 2017, 'An Efficient Secure Communication for Healthcare System using Wearable Devices', *Journal of Computers and Electrical Engineering*, Elsevier, Vol .No 63, October 2017, pp 232-245
25. B. Rodriguez Galdon, C. Tascon Rodriguez, E. Rodriguez Rodriguez, and C. Diaz Romero, "Organic acid contents in onion cultivars (*allium cepa* l.)," *Journal of agricultural and food chemistry*, vol. 56, no. 15, pp. 6512–6519, 2008.
26. J. I. Corcoles, J. F. Ortega, D. Hernandez, and M. A. Moreno, "Estimation of leaf area index in onion (*allium cepa* l.) using an unmanned aerial vehicle," *Biosystems engineering*, vol. 115, no. 1, pp. 31–42, 2013.
27. Vijayakumar, P, Pandiaraja, P, Balamurugan, B & Karuppiah, M 2019, 'A Novel Performance enhancing Task Scheduling Algorithm for Cloud based E-Health Environment', *International Journal of E-Health and Medical Communications (IJEHMC)*, Vol 10, Issue 2, pp 102-117
28. [30] T. Zhao, B. Stark, Y. Chen, A. Ray, and D. Doll, "More reliable crop water stress quantification using small unmanned aerial systems (suas)," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 409–414, 2016.
29. N Deepa, P. Pandiaraja, 2020, "Hybrid Context Aware Recommendation System for E-Health Care by merkle hash tree from cloud using evolutionary algorithm", *Journal of Soft Computing*, Springer, Volume 24, Issue 10, Pages 7149–7161.
30. N Deepa, P. Pandiaraja, 2020, "E health care data privacy preserving efficient file retrieval from the cloud service provider using attribute based file encryption", *Journal of Ambient Intelligence and Humanized Computing*, Springer, <https://doi.org/10.1007/s12652-020-01911-5>.
31. T. Zhao, B. Stark, Y. Chen, A. L. Ray, and D. Doll, "A detailed field study of direct correlations between ground truth crop water stress and normalized difference vegetation index (ndvi) from small unmanned aerial system (suas)," in *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2015, pp. 520–525.
32. K Sumathi, P Pandiaraja 2019, "Dynamic alternate buffer switching and congestion control in wireless multimedia sensor networks", *Journal of Peer-to-Peer Networking and Applications*, Springer, Volume 13, Issue 6, Pages 2001-2010
33. [35] W. Wang, C. Li, and R. D. Gitaitis, "Optical properties of healthy and diseased onion tissues in the visible and near-infrared spectral region," *Transactions of the ASABE*, vol. 57, no. 6, pp. 1771–1782, 2014.
- I. Jolliffe, *Principal component analysis*. Springer, 2011.
34. J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.
35. F. Ellis, "Private trade and public role in staple food marketing: The case of rice in indonesia," *Food Policy*, vol. 18, no. 5, pp. 428–438, 1993.
36. B. Trotter, *Applying price analysis to marketing systems: methods and examples from the Indonesian rice market*. Natural Resources Institute, 1992, vol. 3.
37. P. Pandiaraja, N Deepa 2019, "A Novel Data Privacy-Preserving Protocol for Multi-data Users by using genetic algorithm", *Journal of Soft Computing*, Springer, Volume 23, Issue 18, Pages 8539-8553.
38. E. J. Hannan and J. Rissanen, "Recursive estimation of mixed autoregressive-moving average order," *Biometrika*, vol. 69, no. 1, pp. 81–94, 1982.
39. S. Mahida and B. Patel, "A review of the application of data mining techniques for vegetable price prediction," 2018.
40. Y. Lu, L. Yuping, L. Weihong, S. Qidao, L. Yanqun, and Q. Xiaoli, "Vegetable price prediction based on pso-bp neural network," in *2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA)*. IEEE, 2015, pp. 1093–1096.
41. Z. Su, "Price simulation and prediction of leaf vegetable based on gene expression algorithm," in *IOP Conference Series: Earth and Environmental Science*, vol. 185, no. 1. IOP Publishing, 2018, p. 012013.
42. [45] R. J. Hyndman, Y. Khandakar et al., *Automatic time series for forecasting: the forecast package for R*. Monash University, Department of Econometrics and Business Statistics . . . , 2007, no. 6/07.
43. V. Gomez, "Automatic model identification in the presence of missing observations and outliers. Ministerio de Economía y Hacienda, Direccion General de Analisis y . . . , 1998."
44. L.-M. Liu, "Identification of seasonal arima models using a filtering method," *Communications in Statistics-Theory and Methods*, vol. 18, no. 6, pp. 2279–2288, 1989.
45. G. Melard and J.-M. Pasteels, "Automatic arima modeling including interventions, using time series expert software," *International Journal of Forecasting*, vol. 16, no. 4, pp. 497–508, 2000.

46. J. E. Mays, J. B. Birch, and B. Alden Starnes, "Model robust regression: combining parametric, nonparametric, and semiparametric methods," *Journal of Nonparametric Statistics*, vol. 13, no. 2, pp. 245–277, 2001.
47. N. Balakrishnan, *Methods and applications of statistics in the life and health sciences*. John Wiley & Sons, 2010.