International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Crop Selection Based on Economic Factors: A Review

Vishal Kasa[1], Karan Mahajan[2], Shweta Dhautre[3], Suresh Kapare[4]

[1,2,3]*Student, Dept. of Computer Science Engineering, MIT School of Engineering, Maharashtra, India*
[4]*Professor, Dept. of Computer Science Engineering, MIT School of Engineering, Maharashtra, India*

Abstract - *Farming is a significant occupation in India. It is the primary and major contributor to the Indian economy. Normally, about 70% of the families and 10% of the urban areas are dependent on agriculture. These numbers are decreasing and one of the major reasons is that the farmers are not generating enough revenue to sustain themselves let alone their farms. Farmers are not generating revenue because they are unaware of the economic factors. Also, at times improper planning results in lesser revenue, for example, harvesting at the time when there is a lesser demand. Hence it becomes very important for the farmer to analyze the economic factors while planning the crop. In this study, we have taken a wide view of the exploration done by different researchers satisfying the goal of crop selection dependent on different boundaries that influence the existence of crops. Our research demonstrated that the serious issue confronted while making such a system is that the real data shows sudden bursts and extreme fluctuations which are difficult to predict.*

Key Words: Price Prediction, Crop Selection, Economic Factors, ML, Statistical Methods

1.INTRODUCTION

In India, many farmers suffer huge losses as they are unable to achieve maximum potential profits of their yield. Many times, they are even vastly underpaid. Producers usually have to accept established prices as fact. Their entire livelihood depends on the harvest and with such exploitation, it becomes difficult for the farmers to sustain themselves. FAO has recorded a steady decline of the contribution of agriculture to the GDP of India. While there are many reasons, one of the reasons is the losses incurred by the farmers.

The farmers face these issues as they are unaware of the economic conditions. To minimize these risks, it is very important to select crops that will provide good returns upon harvest. While, majority of the farmers follow traditional cropping plans, it is important to plan crops based on the economic conditions.

In this manuscript we are trying to study and review the research done in the field of crop selection to maximize profit opportunities for the farmer. We are trying to study the different parameters which affect the price of the crops, the approach or strategy used by the researchers and the challenges faced by the researchers.

2. Review Plan

The objective of this study is to find:

- Factors That affect the price of the crop and can these factors be modelled or not.
- The approaches used by the researchers to handle price prediction and their advantages.
- Challenges handled by the researcher to understand the fluctuations in price of the crop.

For this research, we looked through the information bases of Research Gate, ScienceDirect, IEEEXPlore, SpringerLink and Google Scholar, physically, by attempting various words, for example, "price predictions", "economic factors", "crop selection", and so on.

We have limited our research till price prediction where the output of the system is crop price or range of price which is further used as an important feature in crop selection.

The Research consists of all types of manuscripts (Journal, Conference, Book Section) that have referred or gone over in the previously mentioned archives that identify with our research.

For such manuscripts, we have extracted the different kinds of systems they used along with some valid points which encourages us in distinguishing the solution to our research questions.

3. Literature Review

One [1] of the main factors for market instability is fluctuations in market arrivals. As mentioned in the manuscript, analysis of price can overcome this instability. A study of monthly prices of maize in markets of Rajasthan for a period of 12 years (2002-2013) was made. This study intended to develop relevant forecasting models of maize price Nimbahera market of Chittorgarh district of Rajasthan. To choose the best model, the least values of computing measures were considered. MAD, MSE, and MAPE. From the various forms of ARIMA, ANN, and Exponential Smoothing Models. According to the research it was perceived that MAD (79.42), MSE (14461.01), and MAPE (6.62) were least for ARIMA (1, 1, 1) model as compared with the rest models and hence the ARIMA model was considered as the fit model for price forecasting of maize.

One of the most products facing price fluctuation in Turkey are potatoes. So, this study [2] is formed to predict the effect of fluctuations in potato prices on producer income. The sample size was decided using the Neyman Method. The sample size decided to be 56, with a 5% error margin and a 95% reliability limit. The typical lag time was calculated to be approximately 1 year by using the koyck. the particular price for the year 2012 resulted in an income loss of 11,198.6 $/ha. The net profit (Gross Product Value - Production Cost), gross profit margin (Gross Product Value - Variable Cost), cost, the variable cost, and gross production value (production value × price) were calculated. Were calculated using primary data for the research (based on the year 2012) and Secondary data were obtained for the years 2002 - 2012 in reference to the utilization of cultivation areas, production quantity, yield, and potato price. Projection analysis was wont to estimate future prices by using the Quadratic Trend Model. The Koyck model seeks to elucidate changes within the production of a product in terms of the lagged prices for that product, and Schwartz's criterion didn't determine the degree within the lags. The difference between the estimated potato price and therefore the actual price in 2012 led to a producer income loss of 11,198.6 $/ha.

In this manuscript [3], the primary goal is to develop a system that will predict the yield and price using different decision planning models. They used comparative models which is related to varying coefficient regression model. It has 11 predictors. They also used clusters because they are dealing with a small number of similar patterns. There estimation stated that the mixture of different factors is not sensitive to the number of clusters. In this paper they cleared that for the above prediction design planning plays an important role in agriculture, they stated that it is very important to know which crop to produce, how much land to allocate to each crop, when to grow, harvest and sell the crop. There are also uncertain factors like weather and demand so by having limited resources which is used to cultivate ,store and supply crops make the crop decision difficult hence the yield and price will also be unstable, they also include a price prediction confidence band which can help farmer to estimate the maximum loss and gain, they forecasting models are useful when there is rigorous decision planning.

In this manuscript[4], the author says that the vegetable industry plays an important role in providing fresh agricultural products and it's an important source of income but vegetable prices are unstable and change fast so it's very difficult to predict the price of vegetables. For this kind of complex prediction, the author established an integrated prediction model for prediction of price of agricultural production with reference with accurate prediction. Artificial neural networks can deal with such problems in which the background information is not clear and rules establishment is uncertain, it can also achieve nonlinear mapping from input to output. In this paper they used the BP neural network which is a kind of ANN used in weight adjustment and used a backpropagation learning algorithm. After applying all algorithms like BP neural network model, the neural network model based on genetic algorithm, RBF neural network model. The result is the BPNN is good at simulation but relatively poor at prediction, the simulation ability of neural network based on GA is worse than bppn but generalization ability is good, predicting accuracy is better than bppn The RBFN has a similar simulation capability with neural network based on GA. Its forecasting result is better than BPNN, and worse than a neural network based on GA. The integrated prediction model utilizes the advantage, which is provided by a single predicting method, obtains the best prediction accuracy. It can improve the performance of agricultural market price prediction effectively.

In this paper [5], the authors have tested two ML models viz. Radial Basis Function (RBF) and Back Propagation Neural Network (BPNN) for the purpose of price prediction. In this, they have only focused only on one crop of tomato. They claim that even though the prices are affected by various factors, it is not possible to collect data on these factors. Hence, they have only chosen the price of tomato as a function of week for their dataset. They have acquired the data for tomato prices for the district of Coimbatore from January 2009 to March 2012. On which they performed data normalization using Minimax normalization. Then they trained the model on this normalized data and achieved an accuracy of 77.42% for BPNN and 85.55% for RBF.

In this manuscript [6], The author took the data of wholesale price series of soybean and rapeseed-mustard and summarized the price prediction and its fluctuations by using different approaches like ANN, TDNN, The arima model etc. They used two monthly wholesale price series where one series is linear and another was nonlinear in nature, They made a system in which first they gave price series data to arima and then they apply non linearly test and TDNN to predict price. They found out that the nonlinearly in series plays an important role in the accuracy of arima and TDNN model if we are comparing RMSE of price. TDNN has performed better than linear model in the terms of prediction of price fluctuations. The TDNN model also has better accuracy in terms of RMSE and Mad values as compared to arima model. The Results with rape seed mustard data which has a nonlinear pattern is pretty good. They can also improve forecasting accuracy by combining all models together.

In this manuscript [7], for the prediction of the price, the authors have used only the past values of the prices for the crops and employed it on different algorithms such as ARIMA, PLS, ANN and RSMPLS. Based on their study, the

authors also claim that different models are suited for different crops. But overall, ANN and PLS are recommended algorithms. The authors were able to model the system for 100 crops and 15 different regions in Taiwan. And they want to make use of other factors such as climate, cultivation area and location of the market to be considered while predicting the prices.

In this manuscript [8], the authors say that, since the prices of agricultural commodities are very volatile in nature, complex forecast methods may not be appropriate for such scenarios. Instead a ML model can be used to overcome the drawbacks of forecast methods. They have used an MVRVM model on the time series data of the prices of agricultural commodities to predict the prices after different time steps such as 1 month, 2 month and 3 months. MVRVM is a regression tool extension of the RVM model to produce multivariate outputs. They have also compared this model with a feedforward ANN based on the metrics such as RMSPE and RMSE and MVRVM outperforms the ANN model.

In this manuscript [9], the authors have proposed a system for selecting the optimal model for agricultural price prediction. According to their research they claim that different commodities have different distributions (or nature of the graph) and hence there cannot be a single model that works for all agricultural commodities. They have used only price as a function of time for the basis for future prediction, but they have extracted various features such as entropy, curvature, spikes, linearity, etc. from this graph as features to predict the price. They have used algorithms like ANN, ELM and SVR and then classified these systems according to commodity types they are most suitable to also based on a forecast horizon or the time step for which you want the prediction.

In this paper [10], the authors have tried to model the price of a crop based on demand and supply. They have suggested that the price follows a chaotic pattern in which long term boundaries can be predicted but short-term predictions are sensitive and can fluctuate a lot making it difficult to predict short term values. They have presented a model which can generate patterns as visible in real time. This model can handle multiple producers. This model takes into account the abrupt irregular behavior of the market. They wanted to extensively test this model on real time agricultural prices.

4. Observations

By reviewing the manuscripts as mentioned in Section 3, Literature Review, we have noted a few observations which we have described below.

Table -1: Observations

	System used	Parameters	Metric
1	ARIMA (Box Jenkins Model) ANN, Exponential Models (Single, double, and triple)	Past Price	ARIMA Standard Error - 0.69 ANN R Square Error - 0.90
2	Neyman Method, The Koyck model, Cobweb theory, Schwartz's criterion, projection analysis	Current Price, Current Price Simple Index, Consumer Price Index, Real Price, Real Price Index, Estimated Price, Cost and income items, GPV, Variable costs, Production costs, Gross profit, Net profit.	Koyck Model R square Error - 0.80
3	Extended Linear Regression	Temperature, Past Price, Rainfall	Comparative model R square Error - 0.80
4	BP neural network model, the neural network model based on genetic algorithm, RBF neural network model, An integrated prediction model based on the three models above	Past Price	BPNN Accuracy - 75%, Neural network based on GA Accuracy - 75%, RBFN Accuracy - 66.7%, Integrated prediction model Accuracy - 41.6%
5	RBF and BPNN	Past Price	BPNN Accuracy - 77.42% RBF Accuracy - 85.55%

6	ANN, TDNN, ARIMA	Past Price	TDNN RMSE - 0.0278 TDNN MAD - 0.0087
7	ARIMA, PLS, ANN and RSMPLS	Past Price	ANN Accuracy - 87% PLS Accuracy - 75%
8	MVRM	Time series data of price	MVRM Bayesian confidence interval - 0.90
9	ANN, ELM and SVR	Past price, entropy, curvature, spikes, linearity, etc.	SVR MAPE - 8.667
10	Mathematical Model	Past Price	-

4.1 Nature of the Prices of Agricultural commodities

The prices of agricultural commodities are very unstable. It is very difficult to model the prices of these commodities. While it follows a time series model, some papers [1,2,10] claim that the price also follows a chaotic model.

4.2 Factors affecting the Prices

The prices of agricultural commodities or crops are affected by a multitude of factors, including but not limited to Economic factors, Geographical Factors, Social Factors, etc. Many of these features cannot be modelled. But including these parameters to model the price of a crop is a very complex process. Hence many of the researchers have tried to find a direct relation between the price and time and skipping the inclusion of complex parameters.

4.3 Discussion on different systems

Researchers have broadly used statistical methods or machine learning methods to model the behaviour of prices. Two of the most common methods are ANN and ARIMA methods for modelling the price. Some manuscripts have claimed that a single method might not work to model different crops, and hence multiple models were generated for different crops.

These researchers have used different metrics to compute the performance of their models. And hence it has become difficult to compare these models against each other.

But statistical models are better to model the fluctuations in the price. One more characteristic of these models is that as we require a prediction for a smaller and smaller time frame, the accuracy decreases.

5. Abbreviations

Abbreviations	
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Network
BP	Backpropagation
RBF	Radial Basis Function
GA	Genetic Algorithm
TDNN	Time Delay Neural Network
RMSE	Root Mean Square Error
ELM	Extreme Learning Machine
SVR	Support Vector Regressor
ML	Machine Learning
FAO	Food and Agriculture Organization
GDP	Gross domestic product
MVRVM	Multivariate Relevance vector machine
RVM	Relevance vector machine
RMSPE	Root Mean Square Percentage Error
RSMPLS	Response Surface Methodology Partial Least Regression
PLS	Partial Least Regression
BPNN	Backpropagation neural network

6. Conclusion

Modelling the price is a very complex subject as it is affected by numerous simple as well as complex factors. These factors are very difficult to model. After studying the manuscripts, we found out that most of the researchers have used ML based approaches. But using individual models didn't give good results and hence they've used ensemble techniques or different models for different crops.

So, ensemble techniques can be very helpful for price prediction.

With this research, we have likewise brought the exploration done in this field in one spot, so it gets simpler for others to begin their research. We also plan to research further in this topic and create a price prediction system for agricultural commodities using ML methods.

ACKNOWLEDGEMENT

We would like to express our appreciation for providing us with an opportunity to work on this project with **TechMahindra Maker's Lab.** We would also like to thank **Ms. Kanchan Blonde** and **Mr. Saket Apte** for guiding us through this project.

REFERENCES

[1] Hemant Sharma, S.S. Burark; Accuracy Of Different Price Forecasting Models For Maize In Nimbahera Market Of Rajasthan.
[2] Hasan Arısoy, Zeki Bayramoğlu; Determination of the Effect of Price Fluctuations on Producer Income – The Case of Potatoes.
[3] Nantachai Kantanantha, Nicoleta Serban, and Paul Griffin; Yield and Price Forecasting for Stochastic Crop Decision Planning.
[4] Changshou Luo, Qingfeng Wei, Liying Zhou, Junfeng Zhang, and Sufen Sun; Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm.
[5] N. Hemageetha, G. M. Nasira; Radial Basis Function Model for Vegetable Price Prediction.
[6] Girish K. Jha and Kanchan Sinha; Agricultural Price Forecasting Using Neural Network Model: AnInnovative Information Delivery System.
[7] Yung-Hsing Peng, Chin-Shun Hsu, and Po-Chuang Huang; Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets.
[8] Andres M. Ticlavilca, Dillon M. Feuz, and Mac McKee; Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning RegressionRegression.
[9] Dabin Zhang, Shanying Chen, Liwen Ling, And Qiang Xia; Forecasting Agricultural Commodity Prices UsingModel Selection Framework With Time SeriesFeatures and Forecast Horizons.
[10] Zsuzsanna Bacsi; Modelling Chaotic Behaviour in Agricultural Prices Using a Discrete Deterministic Nonlinear Price Model

BIOGRAPHIES

Mr. Vishal Kasa is pursuing BTech (CSE) from MIT School of Engineering, Pune and presently is in Final Year. His Research Interest are Machine Learning, Deep Learning and Data Science.

Mr. Karan Mahajan, completing my Final year from MIT-ADT University, Pune in the field of B.Tech Computer Science Engineering. My Particular field of Interest are in Data Science and Machine Learning

I'm third a year student at MIT ADT University, Pune. Pursuing my B. Tech Computer Science and Engineering. I'm looking forward to opportunities in Software Engineering and Web Development.

Prof. Suresh Kapare, M.E. (Computer Engineering) from Savitribai Phule Pune University. My research area is machine learning. Intended towards solving real life problems of Indian infrastructure of living.