International Journal of Research in Economics and Social Sciences (IJRESS)
Available online at: http://euroasiapub.org
Vol. 7 Issue 5, May- 2017, pp. 20~29
ISSN(o): 2249-7382 | Impact Factor: 6.939 | Thomson Reuters Researcher ID: L-5236-2015

Price forecasting for mono crop pattern using Data Mining Approach

Meenakshi Malik[1]
National Research Centre for Integrated Pest Management, New Delhi-110012

Mamta Bansal[2]
Shobhit University, Meerut-250110

R.P. Agarwal[3]
Shobhit University, Meerut-250110

A. K. Kanojia[4]
National Research Centre for Integrated Pest Management, New Delhi-110012

Mukesh Sehgal[5]
\ National Research Centre for Integrated Pest Management, New Delhi-110012

Abstract- In most parts of India, agriculture has become a risky and farmers commits suicide due to poor yield. The risk is mainly due to availability of water for cultivation and getting profitable prices in market. Prices alter between very high and very low, so it becomes very important for farmers to do wise crop planting to become profitable. Data mining techniques can help to understand the under laying patterns from mass data and if this patterns can be used to help farmers for crop planning, it would reduce the risk and guarantee a minimum profit for farmers to sustain their livelihood. Use of multi model predictor for prediction is a new topic and previously only single classifiers are used. The paper will use multi classifier to create prediction model for agricultural domain in this research project. This work is an attempt in that direction to help farmers in crop planning.

Keywords- Agriculture, Classification, Crop planning, Data mining, Prediction, Water harvesting.
Short running title: Price forecasting for mono crop pattern

INTRODUCTION:

Most of agricultural activities in India are dependent on Monsoon. sharmistha121@gmail.comMonsoon waters are collected in various dams and lakes and used for agriculture. Different crops have different irrigation requirements and water consumption is different for different crops. Many times farmers have planted more water demanding crop but in

case of rain failure they are in total loss. Also the crops depend on soil type and climate conditions. Because the country is quite big, the farmers in one part do not know about the crops cultivated in another area and if they plant the same the supply of that particular crop increases and price cripples. This has been observed in most markets in India. Prices soar some times and it also drops to very low value most times. The main problem for farmers is this price fluctuation. Only if a minimum price is assured for their produce, farming will be profitable. Many agricultural produce like tomatoes where thrown on streets sine price is very low not even able to manage the transportation cost for moving to nearby markets. Many consumable products in other industries have very low fluctuation in prices, but agriculture industries have very high fluctuation and most farmers are confused about which crop to plant for the next season.

The choice of crop to plant is based on multi parameters like water availability, predicted price, climate conditions, soil type, fertilizer availability etc. Climate parameters can be controlled using shade net. Fertilizer availability can be controlled using stocking. Soil type can be managed using pit soils. But the major uncertainty is water availability and the predicted price. Once we these price fluctuation, we see that it has a pattern and the monsoon rain follows a 7-year repletion cycle in most parts of INDIA. So this motivated us to view the crop suggestion as data mining problem and design a predictor model applying data mining algorithms to solve it. Many data mining models have been proposed in agriculture for different applications like rain prediction, crop price prediction but in this work we combine multiple models to suggest the crop to be planted profitably for farmers. Multi-model prediction is being recently used in many fields like stock market price prediction. Multi-model prediction gives far better accuracy then single models. In this work we have also cascaded multi-model predictor together to get the combined multi predictor to predict price of each crop and suggest the crop which is less risky for the farmers.

RELATED WORK

In this section we survey the data mining models used in agricultural domain.

In [1], authors applied image processing and decision system based on Bayesian and Fuzzy K means to detect the weeds avena sterilis which is distributed unevenly in the fields. Based on the detection, tractors were sent to the detected places for weed removal.

In [2], authors applied Multi linear regression and k-means clustering to predict the crop yield. This prediction is based on past year data of rain fall. They modeled the relation between rain fall and production and applying clustering and regression they we able to predict the crop yield for a particular season.

In [3], author's fined tuned C4.5 decision tree algorithm to predict soil fertility based on PH of the soil and the level of micro nutrients minerals in the soil. They decision tree was built to classify the soil fertility to levels of very low, low, moderate ,moderately high , high, very high based on PH and 8 minerals.

In [4], authors evaluated the concept of multi dimensional regression for yield production.

They used MLP (multi-layer Perceptron), RBF (Radial Bias Network), Reg Tree, Support Vector Regression (SVR) to predict the yield. For the datasets they worked, support vector regression was able to give lower RMSE value.

In [5], the level of nitrogen fertilization was predicted used neural networks. The neural network was trained for level of nitrogen fertilization based on electrical conductivity, historical yield, and canopy reflected index. They tested it on winter wheat field on eastern German farm.

In [6], authors applied neural network to predict the vegetable price. They trained the neural network based on past vegetable price and were able to predict the future price. But single variate models like this has low accuracy and does not apply for all types of vegetables.

In [7], authors applied neural network model to predict the vegetable price. After experimenting with BPN (back propagation network), GA (Genetic Algorithm) and RBNF (Radial Bias Neural Network) they proved that individual models accuracy was worst and when they combined multiple models like ANN with GA the accuracy was better.

In [8], authors applied ARIMA model to predict the demand for vegetables in market. They used past demand and built a ARMIA model to predict the demand. However their model worked only for certain vegetables.

In [9], authors developed a decision tree model to predict the fluctuation is pork price in china market. They took 10 year of data and based on it developed a model to predict if price will fall, rise or stable. The model was used an early warning system to alert for pig price risk.

In [10], authors combined PSO (Particle Swarm Optimization) with SVM (support vector machine) to forecast agricultural water consumption. If the water consumption can be modeled, it will become easy to select crops for planting. But crop selection must not be done based on water consumption forecast alone, it must also involve expected price turning profitably for farmers.

In [11], authors propose multi-dimensional model for analyzing land physical properties. They captured multiple attributes of soil and represented in OLAP data cube. The result OLAP data model was used to query for land properties.

In [12] , authors analyzed the data of rural labor, arable land area and the gross output value of agriculture about 30 cities of China based on the decision tree, and adopted clustering analysis method to discretize continuous data during the process of data mining in order to subjectivity comparing to the traditional classification methods. Finally, they generating the decision tree for agriculture, in terms of spatial classification rules and analyzing the rules.

In [13], authors applied data mining model to analyze the droughts. Based on the concept of the new drought index called Vegetation Drought Response Index (VegDRI) using data-mining technology, an Integrated Surface Drought Index (ISDI) was proposed in this work. ISDI improved the original model, adding remote sensed temperature information into the input factors. This index attempt to describe drought from a more comprehensive perspective, the integrated information including: traditional meteorological data, satellite-derived earth surface water and heat environments, vegetation conditions, and inherent properties of the earth's surface. The

Cross-validation results indicated that ISDI construction models for three phases of growth season have very high regression accuracy. The practical application of ISDI in mid-eastern China during the reported dry year 2009 also demonstrated that it can provide accurate and detailed drought condition both at regional and local scale. This investigation showed that ISDI has good application potential for drought monitoring across China.

In [14], authors applied decision tree to analyze the impact of climate factors on soya bean yield. Decision tree model was built to predict the soya bean yield from relative humidity, temperature, and rain fall and evaporation data. But the result is very qualitative and our work tries to propose a quantitative model.

PROBLEM DEFINITION

Given the past N time period values for lake levels and the crop prices in the market, the objective of this work is to predict the lake level and the crop price in N+1 time period and based on the predicted lake level in N+1 time period, select the crop which gives predicted price above a user defined threshold price Tp and sort them in decreasing order of prices.

- Lake levels are chief indicator for farmers to plan their agricultural crops.
- Lake levels exhibiting periodicity property can be predicted for next season for lake levels.
- We collect lake levels for last 10 years and use it predict the lake level for next year.
- So for this we will use three models to learn from data and combine the result of the three models to predict the lake level in future.
- Farmers need to know the price prediction for the agricultural produce, so that they can plan to cultivate the corresponding crop to get maximum profit.
- By using the lake level and the product price prediction they can plan the crop to be planted.

PROPOSED APPROACH

Proposed approach consists of one main algorithm i.e. CSA as described below:

Crop Selection Algorithm (CSA)

The architecture of the proposed Crop Selection Algorithm (CSA) is given below.

The solution has following stages:

1. Lake Level Prediction
2. Crop Filtering
3. Market Price Prediction
4. Crop ordering

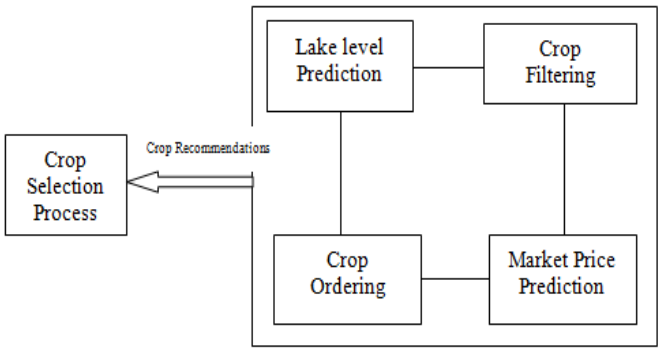

Fig.1: Procedure of Crop selection algorithm

Lake Level prediction module will predict the lake level in future times based on the previous lake level. Crop filtering will invoke select the crops suitable based on the field conditions, climate conditions and lake level. Market price prediction will predict the future price of the selected crops and the crops are ordered in descending order of prices and recommended to the farmers.

The four stages use the following three models

1. ARIMA
2. Multi Perceptron
3. Neural Model

And all the results of the model are combined by using Multi linear regression to predict the crop selection.

a. Lake level prediction

Lake level prediction is a crucial step in our solution. Traditional solutions for lake level prediction are based on using single model but we differ in our solution by employing a multi model. ARIMA and Multi Perceptron are trained to predict the future level and these results are aggregated by using multi layer regression to get the model to predict the lake level.

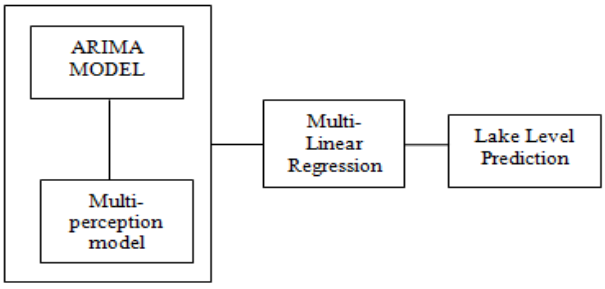

Fig.2: Procedure of Lake level prediction

Say for a future time T, the ARIMA model predicted value is P1 and the multi Perceptron model predicted value is P2; the final lake level prediction i.e. Y is modeled as

$Y = R + b1*P1 + b2*P2$

where, R is the initial level of lake, and b1 & b2 are co-efficient for each model to calculate according to multi-linear regression technique.

b. Crop Filtering

Crop filtering is realized as a decision tree approach. Crops are classified to three levels LOW, MEDIUM, HIGH based on the Lake Level, climate, and soil type. Lake level is transferred to scale of 0 to 100.

Climate is the average temperature during the cropping period and the soil type is sandy, clay. Based on the values of the dependent parameter lake level, climate and soil type the output parameter crop type is determined using a decision tree as shown in table I.

Table I: Values of the output parameter crop type based on dependent parameters like lake level, climate and soil type

Lake Level	Climate	Soil type	Crop Type
<30	>35	sandy	LOW
<30	>35	clay	LOW
<30	27 to 35	sandy	LOW
<30	27 to 35	clay	MEDIUM
30 to 60	>35	sandy	MEDIUM
30 to 60	>35	clay	HIGH
30 to 60	27 to 35	sandy	MEDIUM
30 to 60	27 to 35	clay	HIGH
>60	>35	sandy	MEDIUM
>60	>35	clay	HIGH
>60	27 to 35	sandy	MEDIUM
>60	27 to 35	clay	HIGH

c. Market Price Prediction

Market price prediction is a critical step to determine the profitability. To do the market price prediction we use a multi model. We use ARIMA and Neural network and combine the result using multi linear regression to predict the market price of the crop. Historical prices are crop is collected and used to train the ARIMA and Neural network.

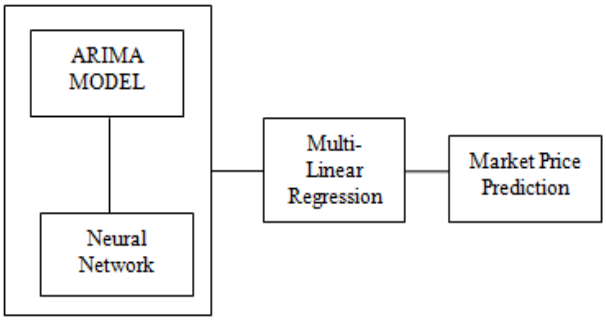

Fig.3: Procedure of Market Price prediction

For each crops in the category provided by crop filtering, market price prediction is applied to predict the market price.

d. Crop Ordering

The selected crops for which the market price prediction is done is then sorted in descending order and recommended to farmers. Farmers can choose the crop from it based on their expertise and profit requirements.

RESULTS

We have collected lake levels in Coimbatore which is located in southern part of India. Choice of Coimbatore is apt since rainfall varies a lot and most of agriculture is dependent on lakes located near to fields. Also price in market is quite volatile. To measure the accuracy of the solution we measured the average profit due to top N recommendations and compared our solution with vegetable price prediction with neural network proposed in [6].

We varied N from 2 to 10 and measured the average profit. From the result we see that average profit is higher in our solution when compared to [6].

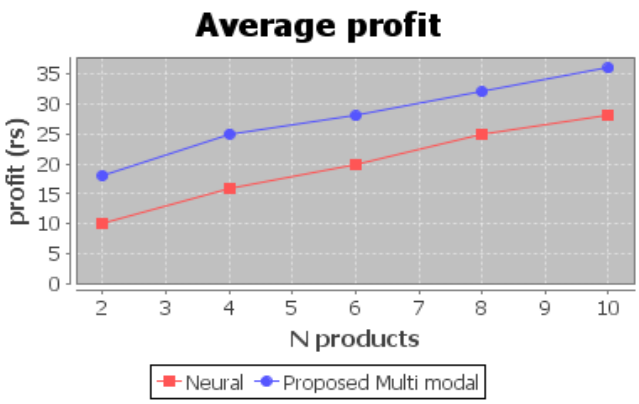

Fig.4: Comparison results of proposed multi model with neural network in terms of profit and number of crop products

We varied the lake level from three ranges <30, 30 to 60, > 60 corresponding to different monsoon levels and for N=5, measured the average profit. From the results we see that our solution adopted fast and was able to give better profit then solution [6].

The result for low lake level of < 30 is below in fig.5

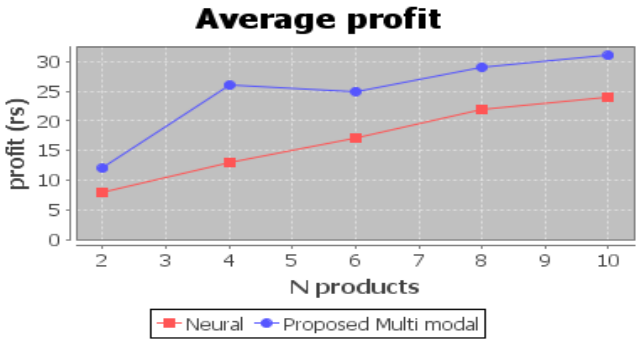

Fig.5: Comparasion results of proposed multi model with neural network for lake level <30

The result for medium lake level of 30 to 60 is below in fig.6

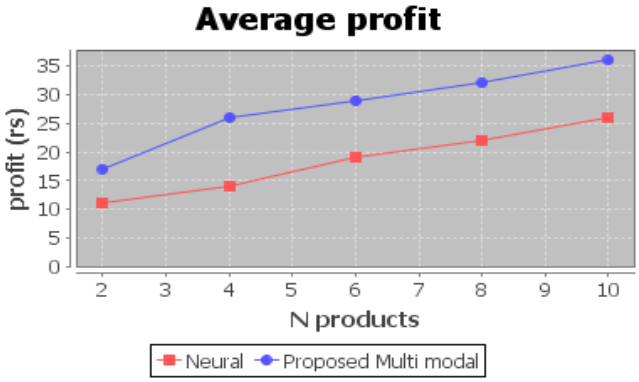

Fig.6: Comparasion results of proposed multi model with neural network for lake level 30-60

The result of high lake level >60 is below in fig.7

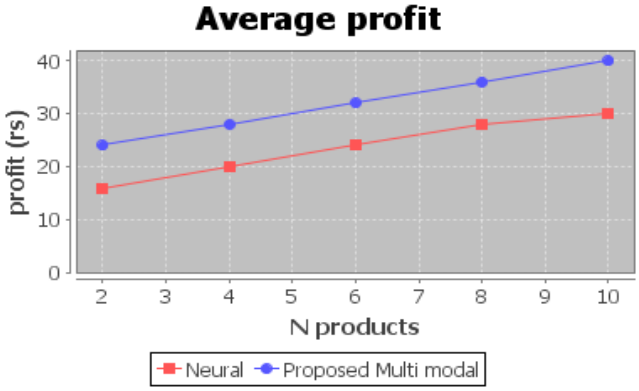

Fig.7: Comparison results of proposed multi model with neural network for lake level >60

CONCLUSION

We have proposed and implemented the proposed crop selection algorithm in this work. Through our experiment, we have proved that our solution is able to select the crops with better profitability. We have implemented this solution for mono crop patterns and in future to reduce the risk and have maximum profitability we can fine tune the solution for diverse crop pattern. By using the multi model paper predicted the product price and compare to actual result and

predicted result with individual models. Overall the research work proved that multi model is able to give better accuracy compared to single model. Also for future work one can try to use others models like Fuzzy etc. to still improve the accuracy.

REFERENCES

[1]Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., & Ribeiro, A. (2007). A vision-based hybrid classifier for weeds detection in precision agriculture through the Bayesian and Fuzzy k-Means paradigms. In Innovations in Hybrid Intelligent Systems (pp. 72- 79). Springer Berlin Heidelberg.

[2]D Ramesh, B Vishnu Vardhan, (2013). Data Mining Techniques and Applications to Agricultural Yield Data. International Journal of Advanced Research in Computer and Communication Engineering 2(9).

[3]Jay Gholap. (2012). Performance tuning of j48 algorithm for prediction of soil fertility. Asian Journal of Computer Science And Information Technology 2: 8 (2012) 251-252.

[4] G Ruß, "Data Mining of Agricultural Yield Data: A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining- Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages : 24-37.

[5]P. Wagner and M. Schneider. Economic benefits of neural network-generated site-specific decision rules for nitrogen fertilization. In J. V. Stafford, editor, Proceedings of the 6th European Conference on Precision Agriculture, 775–782, 2007.

[6]G. M. Nasira, N. Hemageetha, "Vegetable price prediction using data mining classification technique" : Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012

[7]C. Luo Q. Wei L. Zhou J. Zhang and R.S. Sun "Prediction of vegetable price based on Neural Network and Genetic Algorithm" IFIP AICT vol. 346 pp. 672-681. .

[8]M. Shukla and S. Jharjharia "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.

[9]L. Ding, J. Meng, Z. Yang, An Early Warning System of Pork Price in China Based on Decision Tree, IEEE Conference Proceedings, International Conference on E-Product E-Service and E-Entertainment (ICEEE), Henan, China, 1–6, 2010.

[10] S. Lu, Z-J. Cai, X-B. Zhang, Forecasting Agriculture Water Consumption based on PSO and SVM, IEEE Conference Proceedings, 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, 147–150, 2009.

[11] CZ. Radulescu, A Multidimensional Data Model for Analysis of Agricultural Soil Characteristics, IEEE conference on Complexity and Intelligence of the Artificial and Natural Complex Systems, Targu Mures, pp. 126-129, (2008).

[12] G. Yi-yang and R. Nan-ping, Data Mining and Analysis of Our Agriculture Based on the Decision Tree, *ISECS International Colloquium on Computing, Communication, Control and*

Management, Sanya, vol. 8-9, pp. 134-138, August.

[13] Lei Zhou, Jianjun Wu, Jianhui Zhang, Song Leng, Ming Liu, Jie Zhang, Lin Zhao, Fengying Zhang, and Yu Shi, "The Integrated Surface Drought Index (ISDI) as an Indicator for Agricultural Drought Monitoring: Theory, Validation, and Application in Mid-Eastern China", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 6, No. 3, June 2013.

[14] S. Veenadhari and B. Mishra, "Soybean Productivity Modelling using Decision Tree Algorithms", 2011.

[15] Maria Rossana C. de Leon, Eugene Rex L. Jalaok, "A Prediction Model Framework for Crop Yield Prediction", Asia Pacific Industrial Engineering and Management System, 2013.