Crop price prediction in Indian markets using machine learning

Monika T, K Harsha Vamsi, Sowmya G, Indushekar A.C

Project guide: Prof Pooja Agarwal

1 Problem Definition

With advances being made everyday in the fields of machine learning, it is only appropriate to use the ever growing technology to solve a common but rather underrated issue, predicting the future pricing of some of the most popular crops being grown in the Indian states. Farmers everyday face the dilemma of not knowing when the ideal time to sell their crops would be and hence face a lot of problems when it comes to finding the ideal time to make a profit. We plan on solving this problem by helping the farmers find the ideal time to sell their produce.

Key Terms: Crop price prediction, Machine learning

2 Introduction

Agriculture is the supportive backbone of any developing country and hence plays a crucial role in the economic development of a growing nation. But the difficulty of maintaining a profit at all times has hindered the ability of farmers to effectively grow their supply to keep up with the demand.

This is possible by making use of past data available from the AGMarket data hub run by the Indian government. We will use monthly data available for a variety of crops to build our model. There are a lot of advantages to this system:

- Converting this into an application which is reachable to a large number of agriculturalists helps in maximizing the profits.
- Due to the growing digitization and Internet reachability in most rural areas of developing countries, it can work in sound conditions with a variety of parameters.

- Optimizing the system using sound machine learning techniques can result in very accurate and highly efficient outputs.

3 Literature Survey

Price prediction is a very important problem for any farmer as he is the one who should know how much cost he would expect for his crops. In past years, price prediction was done by judging farmer's experience on particular crop and field.

Suppose we have the previous data available in which various corresponding price predictions are recorded and these recorded price predictions are used to classify future price predictions. Price prediction helps the farmers and also Government to make effective decision. Data mining classification techniques can be used to develop an innovative model to predict the market price of respective commodity.

Price prediction is highly useful in agriculture for forecasting, and other involved players manage their inventories effectively and influence their purchasing decisions. Vegetables price is unstable and is changing day by day. Henceforth the prediction of vegetable prices become unpredictable, which leads to the obstacles in promoting the sustained and steady development of the vegetable production.

Market prices of agricultural products are affected by many factors such as climate, supply and demand etc. In the current research, non-linear prediction methods such as neural network and genetic algorithm are used widely, and also make certain results.

To track and to forecast the market prices are both important tasks in agri-management, by which the production schedule can be adjusted to increase the profit. For tracking the crop prices, the Council of Agriculture (COA) establishes an official website that provides open data of daily market prices from over 15 local markets with more than 100 different crops in particular state. Historical climate data and cultivation data on a farmland are of great value for analyzing the potential harvest in the future.

Meanwhile, the trend of crop prices provides important message of the supply and demand, which deeply affects the profit of agri-business. From the perspective of harvest improvement, many previous researches focus on finding the relationship between climate and harvest [1].

At the same time, researches for price forecasting [2] are crucial for production and sales scheduling. In Agriculture pre harvest glut, post harvest loss and intermediary involvement cause producer to get the lowest price in the entire marketing process. The quadratic forecasting model of linear time series compares the prediction results of different time series and different sample data to find the best forecasting model to forecast the price in cites [3].

4 Objective

- Fetch existing data from AGMarkets data hub
- Collect and make dataset with all distinct crop data
- Apply machine learning to fine tune learning model
- Evaluate accuracy and improve by tuning parameters
- Build application to let farmers view state-wise and crop-wise data with predictions for each month with interactive visualizations
- Analyze feasibility to deploy application on large scale

5 Data Description

To be done

6 Methodology

This section talks about the methodology and high level design proposed to achieve the objectives mentioned in the previous section. The approach towards the problem definition is divided into following major components:

- Data Selection : This involves choosing the appropriate data set with minimal noise and also removal of any existing noise present in the data. We will use the AGMarket data hub to collect data per crop and then combine this data to form our own dataset.
- Prediction : To be done
- Performance Analysis : To be done

7 Review of Machine Learning Techniques

Machine Learning techniques can be classified as : supervised and unsupervised learning techniques, also termed as predictive and descriptive methods respectively[15]. The Supervised technique requires a training set of data(with features and labels for which the target feature is known with confidence. An algorithm is trained on this set and the mapping that is obtained as a result of training is applied to other objects for which class label is not available. In case of Unsupervised learning, labelling of data is not done and requires an initial input to one or more of adjustable parameters and the solution depends on the input. Some of the important ML techniques are described below:

- Linear Discriminant Analysis(LDA) : In this method, the classifier tries to find a linear boundary that best separates different classes in the data. This techniques gives the optimal Bayes' classification assuming same covariance for all classes.
- Naïve Bayes' method : Here, classification os an arbitrary number of independent variables and is used when data has many attributes. The data is classified either as categorical or numerical.
- Decision Tree : This method constructs a tree structure for classification or regression purposes. Each node of the tree splits the training set based on a feature. The first or the root node is based on a feature that qualifies to be the best predictor. A splitting criterion or a decision rule splits other nodes into child nodes. This method can handle both categorical and numerical data.

- Random Forest : is an ensemble of multiple decision trees.Each tree is constructed by selecting a random subset of attributes from the dataset. Random Forest techniques works well for large datasets.
- Support Vector Machine(SVM) : This is particularly suitable for binary class discrimination. An optimal hyperplane is obtained that maintains largest minimum distance from the training data.
- Principle Component Analysis(PCA) : This assumes data is real , continuous and normalized. This could fail in case of non-normalized data.

Thus, the choice of the classifier is decided by the nature of the separability of data. It is seen that if data is not linearly separable, the classifier chosen may not yield good results.

8 High Level Design

To be done

9 Proposed Metrics

To be done

10 Model

To be done

11 Assumptions

To be done

12 Outcomes

To be done

13 References

[1] J. Arno, J. R. Rosell, R. Blanco, M. C. Ramos, J. A. MartinezCasasnovas;Spatial variability in grape yield and quality influenced by soil and crop nutrition characteristics; Precision Agriculture, vol. 13, pp. 393-410,2012.

[2] Sarah Abadan and Ani Shabri;Hybrid empirical mode decompositionARIMA for forecasting price of rice; Applied Mathematical SCiences, Vol. 8, pp. 3133-3143,2014.

[3] Y. Chun-Yan, M. Jun and Z. Yu-Yan, "Online Price Extraction and Decision Support for Agricultural Products", Proceedings of the 2009 International Conference on Information Management, Innovation Management and Industrial Engineering , IEEE Computer Society, TBDXi;an, China, vol. 2, no. 36, (2009) December 26-27.

[4] Changshou Luo1, Qingfeng Wei1, Liying Zhou2, Junfeng Zhang1, and Sufen Sun1," Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm".

[5] Manpreet Kaur , Heena Gulati, Harish Kundra, " Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications ".

[6] Yung-Hsing Peng, Chin-Shun Hsu, and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets".