CROP PRICE FORECASTING SYSTEM USING SUPERVISED MACHINE LEARNING ALGORITHMS

Rachana P S[1], Rashmi G[2], Shravani D[3], Shruthi N[4], Seema Kousar R[5]

[1,2,3,4]BE, Department of CSE, NIE Mysore, Karnataka, India
[5]Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India

---***---

Abstract - *In the recent years due to the uncertain climatic variations, natural calamities and other issues the price of the crop has varied to a larger level. Farmers are uninformed of these uncertainties, which causes massive loss. The developed Crop Price Forecasting System is a well-designed System which provides accurate results in predicting price and profit of the crop. The system comprises of two actors, the Administrator and the Agricultural Department. The Admin maintains the entire System. The Departments performs two main roles, the Price Prediction and the Profit Prediction. The Parameters considered for Price Prediction are- Rainfall, Maximum-trade, Minimum Support Price (MSP),Yield. The Parameters considered for Profit Prediction are- Crop Price, Yield, Cultivation Cost, Seed Cost. To predict the Price of the Crop we use Naïve Bayes Algorithm which is a Machine Learning Classification technique. To predict the Profit of the Crop we use K Nearest Neighbor(KNN) which is a Supervised Machine Learning Classification Algorithm. The Department Head can select the Crop and the Prediction methodology and provide the required parameter values. The algorithms run in the background and give the Output to the Department Heads which is in turn conveyed to the Farmers' for better pre-knowledge about the outcome. Thus, the System gives a beforehand prediction to the Farmers' which increases the rate of Profit to them and in turn the country's economy.*

Key Words: **Forecasting System, K- Nearest Neighbour (KNN), Naïve Bayes, Price Prediction, Profit Prediction.**

1. INTRODUCTION

Agriculture is the main pillar of economy in our country. Most of the families rely on Agriculture. Country's Gross Development predominantly lean on Agriculture. 60% of the land is utilized for Agriculture to adequate the requirements of the Country's population. To meet the requirements, modernization in Agricultural practices is required. Thus, heading towards the growth in Farmers' and Country's economy. In the recent years there has been an inconsistency in the prices of multiple crops which in turn has increased the menace encountered by the Farmers. The main purpose of the Forecasting System is to ensure that the Farmers make a better-informed decision and manage the price risk. In the recent times, most of the Farmers are deprived of the knowledge about the various breeds crops, season of sowing seeds, cultivation methodologies, cultivation cost and other conditions. Also, the farm

productivity can be increased by understanding and forecasting crop performance in a variety of environmental conditions. An effective Crop price forecasting system can give out possibilities for customers which can satisfy the customers to a greater context. Finally, the results are displayed for the Department Heads based on Crop chosen and the Parameters supplied.

1.1 Price Prediction System

In the paper the emphasis is on machine learning technique to predict the Price of the Crop using the Naïve Bayes Algorithm. The price of the crop is determined by recognizing the patterns in our training dataset which is given as one of the inputs to the Algorithm. The inputs values for the parameters(Yield, Rainfall, Minimum Support Price, and Maximum Trade) are taken by the user and fed to the algorithm. The other Parameters to the Algorithm are- Probability, New Record Input and number of Dataset Parameters.

1.2 Profit Prediction System

In the paper there is also focus on a machine learning technique to predict the profit of the Crop which is K Nearest Neighbor technique. The profit of the crop is determined so that the farmers are aware about their net profit before sowing the crop. The profit is calculated based on shortlisting the similarities found in the dataset values. The algorithm takes the parameters like yield, price, cultivation costs and actual seed cost.

2. RELATED WORK

[1]The paper aims at analyzing the demand of the crop by the predicting the price of the crop at the correct scenario. The researcher collects the dataset from the Taiwan markets by the knowledge of market prices of over 15 local markets. The price is predicted using the well-known algorithms such as the partial least square (PLS), autoregressive integrated moving average (ARIMA) and the artificial neural network (ANN). The paper compares the performance of these four algorithms with the prices so obtained from the local markets. The crops considered for analyzing are cabbage, watermelon, bok choy and cauliflower. According to the results so obtained the PLS and ANN algorithms gave a

better result than others for both short term and long-term forecasting.

[2] The research aims over choosing an appropriate crop for an area selected by the user thus helping the farmer to take better and wise decisions. It also suggests the rank of the crop based the suitability to that area. Thus, the farmers get to know the compatibility of the chosen crop and the area selected. The research is carried out for 6 crops which are potato, aus rice, boro rice, wheat, aman rice and jute and the datasets of the previous years are collected respectively. The prediction which predicts the output is done by analysing the dataset using supervised machine learning technique such as K nearest neighbour regression algorithm and decision tree learning (ID3).

3. METHODOLOGY

Naïve Bayes Algorithm for Price Prediction

Naïve Bayes Algorithm is a Machine Learning Classification Technique that uses Bayes' Algorithm. The assumptions are independent and unrelated to other features which we use for price prediction.

Steps followed in the Algorithm-

The algorithm has four inputs to it-

1)The number of parameters (m) -The prediction is based on the various parameters which are provided as the inputs to the algorithm through our frontend. These parameters affect the price of the crop as mentioned in the previous sections. The parameters considered are: Rainfall, Minimum Support Price (MSP), Yield and Maximum Trade.

2)The Probability of possible outcomes (p) – There can be any number of possible outcomes which depend on the number of records for each crop in the training dataset.

3)Training Dataset- The previous years' data provided in a structured format (Excel sheet) is given as an input to the algorithm from which the records would be filtered based on the current input parameters of a crop to which the price has to be predicted.

4)Current Record Parameters-A new current year record which contains the values of the parameters for which the possible price must be predicted.

The algorithm works based on the formula-

   P=[n_c+(m*p)]/(n+m)

where

n - The number of previous records where the value of the parameters is as same as the new record parameter value.

n_c- The number of previous records that has the result value equal to the possible outcome value.

P- The probability of a crop for a certain price.

The one value with greater probability would be chosen as the most accurate value.

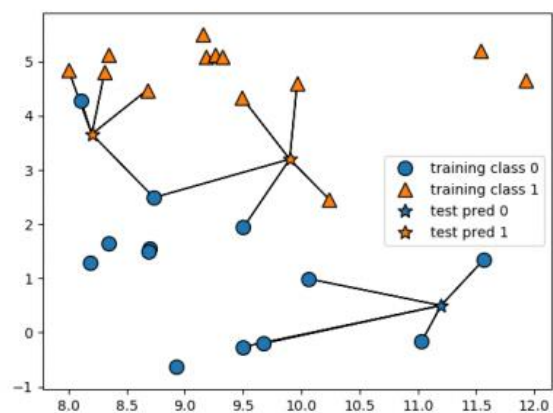### K Nearest Neighbour (KNN) Algorithm for Profit Prediction

Fig -1: Concept involved in KNN

K nearest neighbour is a supervised machine learning classification algorithm which in our case is used for predicting the profit of the crop. This algorithm basically searches for similarities in the entire dataset. Based on the most similar or nearest values the outcome is predicted. This algorithm is chosen over other algorithms because of its higher convergence speed and simplicity. The inputs to the algorithm are:

-The input Parameters

-The value of k

-The training dataset

The Euclidian Formula-

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

Steps involved are:

1) The entire dataset is divided into number of classes or possible outcomes depending upon the number of records in the dataset and the probability of the profit. The algorithm predicts that the input given falls into which one of the classes.

2) We decide the value of k which stands for number of nearest neighbours to be considered. The value of k is decided based on the number of records present in the dataset. The accuracy of the result is based on deciding the appropriate value of k which shouldn't be too high or too low. In our dataset having 100 entries of records, the appropriate value for k is taken as 10.

3) The Euclidian distance is calculated independently for each record in the dataset by taking the difference of the values present in the dataset records and the input values.

4) The k nearest neighbours, that is the k minimum distances are used for predicting the profit value of the crop.

4. FUTURE ENHANMENTS

The future enhancement of our application is to implement this application in every location (by taking a GPS module) as much as possible by extracting the dataset for that location which further increases the accuracy and probability. Implementing the chat portals making the application more user friendly can be another scope for enhancement. Thus, ensuring that the information is available at most ease for them and the more accurate prediction increases the probability of profit for the Farmers.

5. CONCLUSION

The research aims at predicting both the price and profit of the given crop before sowing. This web application runs on efficient machine learning algorithms and technologies having an overall user-friendly interface to the users. The training datasets so obtained provide the enough insights for predicting the appropriate price and demand in the markets. Thus, the system helps the farmers in reducing their difficulties and stop them by attempting suicides.

REFERENCES

[1] "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets", Yung-Hsing Peng, Chin-Shun Hsu, and Po-Chuang Huang, 2017 IEEE

[2] "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques", Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, 2015 IEEE