

# Applying Data Mining Technique to Predict Annual Yield of Major Crops

Praful S. Tayde<sup>1</sup>, Balkrishna K. Patil<sup>2</sup>, Rajesh A. Auti<sup>3</sup>

ME 2<sup>nd</sup> Year, Computer Science & Engineering, Dr. Seema Quadri College of Engineering & Technology, Aurangabad, India<sup>1</sup>  
Assistant Professor, Computer Science & Engineering, Dr. Seema Quadri College of Engineering & Technology, Aurangabad, India<sup>2,3</sup>

prafull.tayde77@gmail.com

**Abstract**— Climate and other environmental changes impact on agricultural economy of any country. Production of crop normally depends on factors like biology, climate, economy and geography, this factors lead to impacts on agriculture. By applying different methodologies and techniques on yields of crops, it is possible to obtain information about which help to government organization for making good decisions and applying different policies, also it's helpful for farmers to make better plan to increase the production.

**Keywords:** - Data mining, crop productivity, crop analysis; yield prediction, K-means

## I INTRODUCTION

Maharashtra is the third largest state in area and second largest state in population of India. It has an area of 307,713 sq. km. with 35 districts, 358 blocks and 43711 villages and a population of 112,372,972. The 45% population of the state is urban.

Agriculture is the mainstay of the state of Maharashtra. Maharashtra's economy is predominantly agrarian. It is the main occupation of the people. Both food crops and cash crops are grown in the state. Principal crops include rice, Jowar, bajra, wheat, pulses, turmeric, onions, cotton, sugarcane and several oil seeds including groundnut, sunflower and soybean. The state has huge areas, under fruit cultivation of which mangoes, bananas, grapes, and oranges are the main ones. [6]

Maharashtra's different district has varying climates and so it is very important to consider environmental factors of these separate areas. This will help to choose the best districts for cultivation of different type of crops. Rainfall also varies from district to district and this has a huge impact on farming because while too little or too much rain can kill crops, the proper amount of rain leads to an ideal crop yield. With rainfall comes humidity and since rainfall varies from district to district so does humidity. Humidity causes changes in the level of water that can be absorbed by atmosphere which can cause crops to remain too wet or too dry and so to get proper yield, a district with an ideal average annual rainfall and humidity is required.

Common specific problem that occurs is yield Prediction. As early into the growing season as possible, a farmer is interested in knowing how much yield he is about to expect. [6]

## II CLIMATIC CONDITIONS

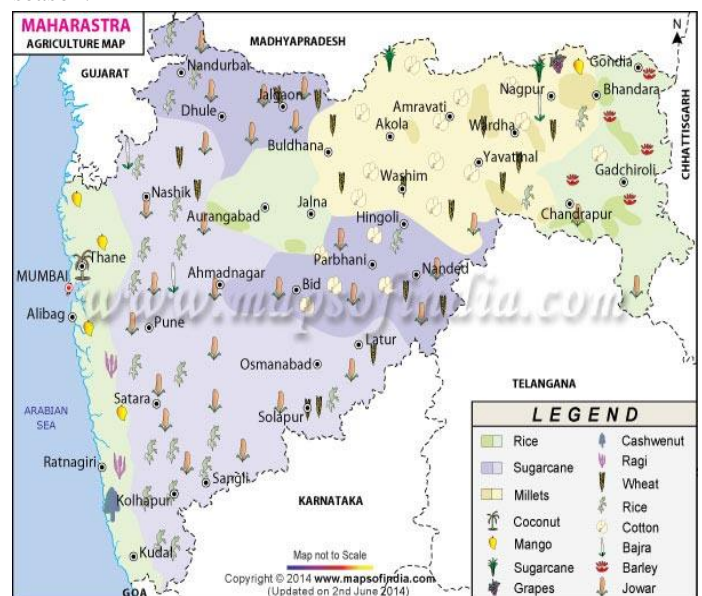
Maharashtra has typical monsoon climate, with hot, rainy and cold weather seasons. Tropical conditions prevail all over the state.

**Summer:** March, April and May are the hottest months.

**Rainy Season:** Rainfall starts normally in the first week of June. July is the wettest month in Maharashtra, while August too gets substantial rain. Monsoon starts its retreat with the coming of September from the state.

**Winter:** Winter is Cool dry spell and pleasant weather prevails from November to February.

Eastern part of Maharashtra sometimes receives some rainfall. Temperature varies between 12°C-34°C during this season.



The total irrigated area which has been used for crop cultivation is 33, 500 square kilometers. The agriculture in state is predominantly rain-fed.

The state has 24 per cent of drought-prone area of the country. However state has potential for growth in agricultural sector in spite of challenges.

### III INPUT VARIABLES

#### The Climatic variables:

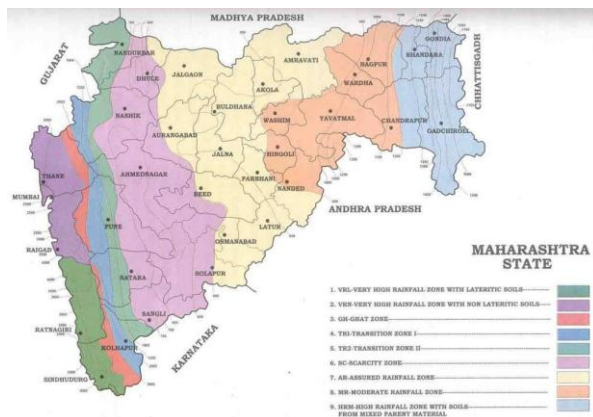
I) Max Temperature: Temperature variation in year puts a great impact in that year's crop production. Hence we consider both the maximum as well as the minimum temperature.

II) Min Temperature: The average yearly minimum Temperature considered in Celsius.

III) Rainfall: The year that has the highest average rainfall leads for maximum crop yield production in that Year. The average yearly rainfall was considered by calculating average from the monthly rainfall (mm) of each district.

IV) Humidity: Average yearly humidity for each district is considered in percentage.

V) Average Sunshine: This attribute was considered in hours as a yearly average for each district. The amount of sunshine received on areas each year greatly effects the production of green crops as it directly affects the photo-synthesis process in plants.



#### b) The biotic input attributes:

i) Max pH: pH value is important because it indicate how acidic soil is. It's scaled is defined by a value of 7, where soil pH above 7 meaning alkaline and below 7 meaning acidic. Crop production is highly affected by the variations of pH in soil.

ii) Min pH: Minimum pH of a district's soil.

iii) Soil Salinity: Soil salinity means the amount or content of salt in soil. It measure in MMHOS/cm, it has ranges (<2), (2-4), (4-8) and (8-15). Salinization process increase salt content. Too high soil salinity can cause a Detrimental effect towards crop production and Yield.

#### c) Central Area:

i) Irrigated Area: Crop production is depend on the actual area of land that has been irrigated throughout the year. Hence irrigated area is considered for the selected districts in hectares.

ii) Cultivated Area: The area that has been used to Cultivate each crop also regulates the amount of Production of the crop. Areas were taken in Hectare unit.

### IV DATA MINING TECHNIQUE

Many techniques for learning rules and relationships Automatically from diverse data sets were developed [1], to simplify the often tedious and error-prone process of acquiring knowledge from empirical data. While these techniques are plausible, theoretically well-founded, and perform well on more or less artificial test data sets, they depend on their ability to make sense of real-world data.

A web based expert information system based on ID3 algorithm was studied [2] in which an expert system provides advisory services to Tomato growers regarding pests, diseases and their control measures. The web based system has also provision for the growers to interact with other growers on the management practices of tomato crop cultivation.

The advanced version of decision-making tree algorithm IBLE that it mainly uses in the information theory [3] The channel capacity concept to take chooses the important characteristic to the entity in the measure. Combines the rule with many characteristics the point to distinguish the example can effectively the correct distinction. They applied this algorithm in the oral cavity disease diagnosis; the experimental result indicated this algorithm has the very strong recognition capability to agriculture case diagnosis to very good assistance diagnosis function. The application of information technology in agriculture accelerates the digitization of agriculture information [4] presented a new improved CA algorithm based on traditional decision tree method. It introduces a pre-treatment theory about double dimension reduction which can deal with large and high-scale datasets. By using CA algorithm in maize seed breeding, they analyzed the potential rules and found out useful Information from it for direct growth of maize. Their experiment showed the improved CA algorithm can obtain more intuitive and efficient information.

Crop forecasts are typically needed between the time of planting and the time of harvest. These associative models use past data to estimate the models and "future" data for prediction. Future data can be implicit or explicit. In general, forecasting methods can be subdivided into two categories: qualitative and quantitative.

Linear Regression: It is a statistical measure that can be used to determine the strength of the relationship between one dependent variable and a series of other changing variables known as independent variables (regular attributes). If independent variable contains multiple input attributes like in our research (rainfall, sunshine hours, humidity, pH etc), then it is termed as multiple linear regressions. Linear regression provides a model for the relationship between a scalar variable

and one or more explanatory variables. This is done by fitting a linear equation to the observed data [5].

Neural Net: An artificial neural network (ANN) is mathematical model or computational model inspired by the structure and functional aspects of biological neural networks for instance in our brains. In most cases an ANN is an adaptive system that modifies its structure based on external or internal information that flows through the network during the learning phase. The basic neural network model consists of three layers: the input layer, the hidden layer and an output layer [5].

#### **V. CONCLUSION**

It is found that by applying different methodologies and techniques on yields of crops, it is possible to predict crop yield information, which helps to government organization for making good decisions and applying different policies, also it's helpful for farmers to make better plan and increase their production, which lead to increasing the country's overall profit.

#### **REFERENCES**

- [1] McQueen Robert J, Garner S.R., Nevill Manning C.G. Ian H. Witten, 1995. "Applying machine learning to agricultural data". Computers and Electronics in Agriculture. Vol. 12:pp. 275-293
- [2] Babu, MSP., Ramana Murthy, NV, SVN Narayana, 2010. "A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms". Int. J. of Comp. Sci., and Information Technologies. Vol. 1(1) . pp. 6-15.
- [3] Yue Jin Hai, Song Kai, 2010. "IBLE Algorithm in agricultural disease diagnosis". In third International Conference on Intelligent Networks and Intelligent Systems held at Shenyang, Liaoning China during November 01- November 03.
- [4] Jianlin Ji Dan, Qiu Chen, Jianping Chen, Li He Peng, 2010. "An improved decision tree algorithm and its application in maize seed breeding". Sixth International Conference on Natural Computation, held at Yantai, Shandon 10-12<sup>th</sup> January. pp. 117-121.
- [5] Ye, Nong; Data Mining: Theories, Algorithms, and Examples, CRC Press, 2013.
- [6] <http://agricoop.nic.in/sites/default/files/Maharashtra>