

ANALYSIS OF CROP YIELD PREDICTON USING DATA MINING TECHNIQUE

Ms. Fathima¹, Ms. Sowmya K², Ms. Sunita Barker³, Dr. Sanjeev Kulkarni⁴

^{1 2 3} Students, Dept. of Information Science and Engineering, Yenepoya Institute of Technology, Moodbidri,
Karnataka, India

⁴ Associate Professor, Dept. of Information Science and Engineering, Yenepoya Institute of Technology, Moodbidri,
Karnataka, India

Abstract – India is generally an agricultural country. Now a days the most important emerging field in the real world is agriculture and it is the main occupation and backbone of our country. Recent developments in Information Technology for agriculture field has become an interesting research area to predict the crop yield. Crop yield prediction is the methodology to predict the yield of the crops using different parameters like rainfall, temperature, fertilizers, pesticides and other atmospheric conditions and parameters. Data Mining techniques is very popular in the area of agriculture. Data mining techniques are used and evaluated in agriculture for estimating the future years crop production. This paper presents a brief analysis of crop yield prediction using K-Nearest Neighbor(KNN) Algorithm for the selected region that is Mangalore, Kasargod, Hassan, Kodagu in India.

Key Words: Data Mining, Yield Prediction, Data sets, K-Nearest Neighbor(KNN) Algorithm.

1. INTRODUCTION

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops.[2]

People of India are practicing Agriculture for years but the results are never satisfying due to various factors that affect the crop yield. To fulfill the needs of around 1.2 billion people, it is very important to have a good yield of crops. Due to factors like soil type, precipitation,

seed quality, lack of technical facilities etc the crop yield is directly influenced. Hence, new technologies are necessary for satisfying the growing need and farmers must work smartly by opting new technologies rather than going for trivial methods.

Data Mining is the process of extract helpful and significant information from huge sets of data. Data Mining in agriculture field is a comparatively novel research field. Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is concerning to be expecting. In the earlier period, yield prediction was performing by considering farmer's experience on particular field and crop. In any of Data Mining actions the training data is to be collected from past data and the gathered data is used in terms of training which has to be exploited to study how to categorize future yield predictions. [7]

2. LITERATURE SURVEY

- [1] This paper describes and gave the details us for list of used methods, In India there are dissimilar agriculture crops production and those crops depends on the several kind of factors such as environmental science, economy and also the geographical factors covering such methodologies and methods on historic yield of dissimilar crops, it is possible to get info or data which can be supportive to farmers and government organizations for creation well decisions and for make better rules which help to increased production.
- [2] This paper discussed a several subdivision in India is facing rigorous problem to make the most of the crop productivity. More than 60 out of a hundred the crop still depends on monsoon

rainfall. Current growths in Information Technology for agriculture field have developed an interesting research area to forecast the crop yield. The problematic of yield prediction is a major problem that remains to be solved based on accessible data. Data mining methods are the better selections for this purpose. Different Data Mining methods are used and evaluated in agriculture for approximating the upcoming year's crop production.

- [3] This paper describes that the suggested Fresh market fruits like apples are graded into quality groups according to their size, color and shape and to the attendance of defects. This paper presents the three former points on the basis of a literature review, the research outcomes being absorbed on the last point: having extracted data from images acquired on fruits, the paper defines a classifying technique which was implemented on an existing machine and tested on Jon gold apples (bi-color fruits).
- [4] This paper presented a Precision agriculture (PA) and information technology (IT) are closely interwoven. The former frequently refers to the application of nowadays' technology to agriculture. This paper deals with suitable modeling methods for those agricultural data where the objective is to uncover the surviving patterns. In specific, the use of feed-forward back propagation neural networks will be evaluated and suitable parameters will be projected.
- [5] This paper discussed that the data mining extraction of unseen predictive information from huge records, is a powerful new technology with great potential to help companies focus on the most significant data in their Data warehouses. Data mining tools predict upcoming trends and performance and growth, allowing businesses to make proactive, knowledge driven decisions. This article gave us a detail project that is smearing a range of machine learning plans to problems in agriculture and horticulture.

2.1 OBJECTIVES

The main objective is collecting agricultural dataset which can be used to analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, advanced methods can be

introduced to predict crop yield and it is also helps the farmer to choose the most suitable crop, thereby improving the value and gain of the farming area.

3. EXISTING SYSTEM

Initially the raw data set was collected and it is subjected to preprocess for noise removing (replacement of missing values) and computational methods. From that dataset, it is subjected to Feature selection for make a predictive modeling. In this proposed approach it is mainly focused on Regression Techniques. Various regression analysis should be performed and it was compared and tested.

Regression analysis is a form of predictive modeling technique which investigates the association between a dependent (target) and independent variable(s) (predictor). This technique is used for forecasting, time series modeling and discovers the causal effect relationship between the variables. Regression analysis indicates the significant relationships between dependent variable and independent variable and it indicates the strength of impact of multiple independent variables on a dependent variable. [6]

4. PROPOSED SYSTEM

In this proposed system, the datasets are collected and refined based on the commonality. The input parameters are given. By analysing and predicting using KNN algorithm, the result are produced and some suggestions are given.

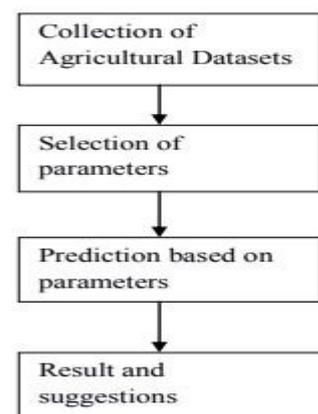


Fig. -1: Proposed Architecture

Step 1:

The datasets have been collected and refined based on commonality uses such as location, crop, Area, soil type, temperature, humidity etc. From these parameters name

of the crop and net yield rate of the crop can be predicted.

Step 2:

Based on various analyses the parameters location, soil type and area are taken as input and prediction have been undertaken. The attribute soil type specifies the type of soil in a particular region such as Coastal alluvials, Laterite soil and Dark brown alayey soil and the attribute location specifies the 4 different areas such as Mangalore, Kodagu, Hassan, Kasargod.

Step 3:

By using KNN algorithm, the particular crop has been analysed and predicted by taking various parameters into an account such as soil type area and location.

Step 4:

By analysing and predicting the crop name and price of particular crop can be found out. This helps the farmers to take the correct decision to sow the crops such that yield rate can be increased.

5. SYSTEM REQUIREMENTS AND SPECIFICATION

System Requirements Specification is a document or set of documentation that describes the features and behaviour of a system or software application. It includes a variety of elements that attempts to define the intended functionality required by the users to satisfy their different users.

5.1 FUNCTIONAL REQUIREMENTS

The Functional Requirements Definition reports and tracks the basic information expected to effectively portray business and handy necessities. The Functional Requirements Definition report is made in the midst of the Planning Phase of the endeavor. Its objective gathering is the endeavor boss, errand gathering, wander bolster, client/customer, and any accomplice whose information/respect into the necessities definitions system is required.

5.2 NON-FUNCTIONAL REQUIREMENTS

A non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. Non functional requirements specifies the quality attribute of a software system. The software system can be judged based on reliability, security, maintainability, performance, portability, scalability and flexibility

5.3 HARDWARE REQUIREMENTS

Computer hardware specifications are technical descriptions of the computer's components and capabilities. Processor speed, model and manufacturer. Processor speed is typically indicated in gigahertz (GHz). The higher the number, the faster the computer. Random Access Memory (RAM). This is typically indicated in gigabytes (GB). The more RAM in a computer the more it can do simultaneously. Hard disk (sometimes called ROM) space. This is typically indicated in gigabytes (GB).

Processor	Intel Pentium/Core - 1.7GHz and above
Memory	1GB and above
Storage	80GB minimum free space
Graphics	1GB and above

5.4 SOFTWARE REQUIREMENTS

Operating System(s)	Windows 7, 8, 10
Programming Language	Python
Framework	Pycharm

6. SYSTEM DESIGN AND IMPLEMENTATION

6.1 UML DIAGRAM

UML diagram is a diagram based on the UML(Unified Modeling Language)with the purpose of visually representing a system along with its main actors ,roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system.

In Fig.2 which explains the UML diagram for our proposed system. Here crop dataset, crop recommender and users. The crop dataset will be separated into 2 datasets:Training dataset and testing dataset.

Training data: A training data set is a set of data used to discover predictive relationships. Training set is used in intelligent systems, machine learning, genetic programming and statistics.

Testing data: Test data is the input given to a software program. It represents data that affects or is affected by the execution of the specific module.

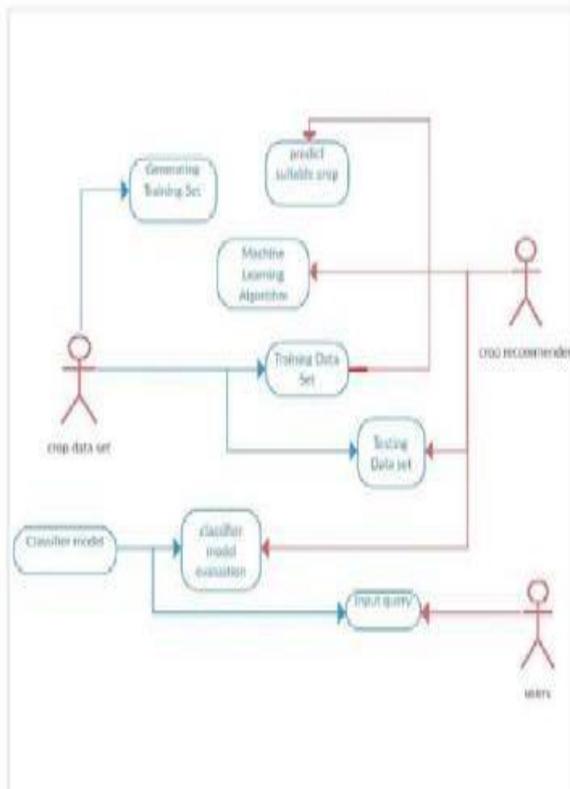


Fig. -2: UML Diagram

This system works as follows:

Step 1: Giving input (training data and testing data) user provide input to system as training data and testing data.

Step 2: compute the prediction: Testing data and training data is given to K-NN algorithm which creates clusters from given data. K-NN algorithm gives relevant data, patterns from given datasets.

Step 3: Data Acquisition: Depending on the current user location, the system mines the soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set.

Step 4: Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop. The prediction is based on past production data of crops i.e.: identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner. Step 5: Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

6.2 SYSTEM ARCHITECTURE

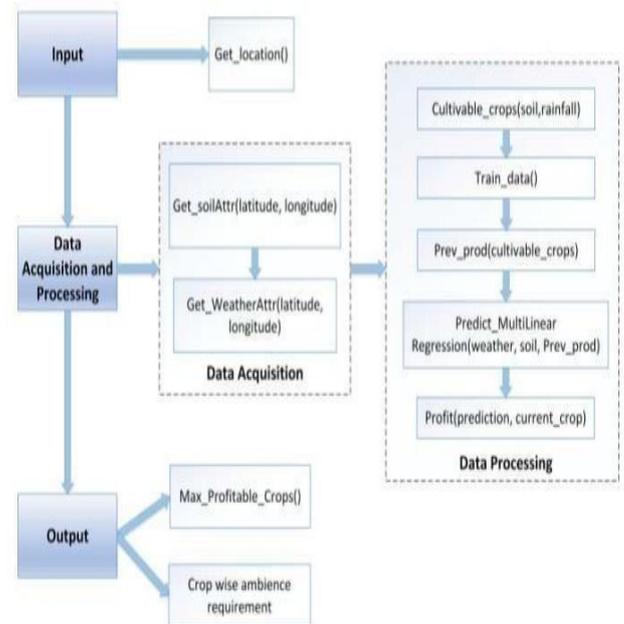


Fig. -3: System Architecture

Input: The prediction of crop is dependent on numerous factors such as Soil Nutrients, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system.

Data Acquisition: Depending on the current user location, the system mines the soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set.

Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop

Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

7. RESULTS AND SNAPSHOTS

This chapter contains the snapshots of the graphical user interface(GUI)of the proposed system showing the interface, intermediate results and final result of crop prediction.

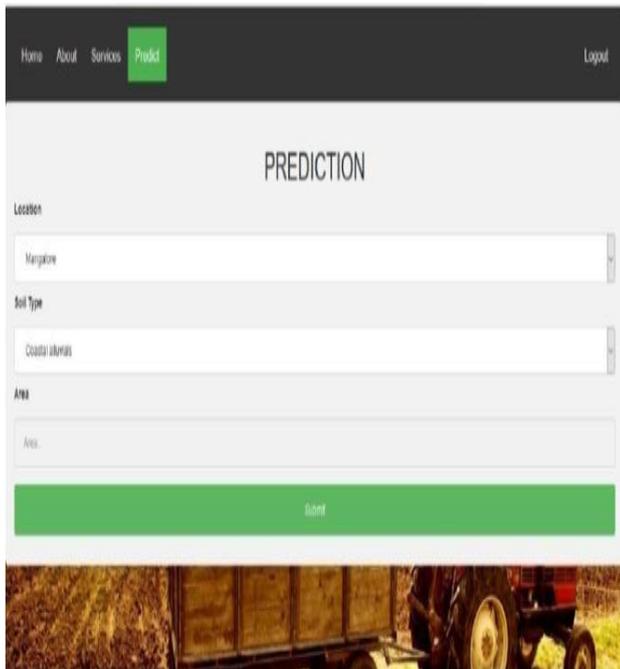


Fig. -4: Prediction Page

The Fig. 4 shows the Prediction page. In this page user has to enter a value to the field such as location, soil type and area. Once user fills all the inputs the user has to click submit button then the result will be print.

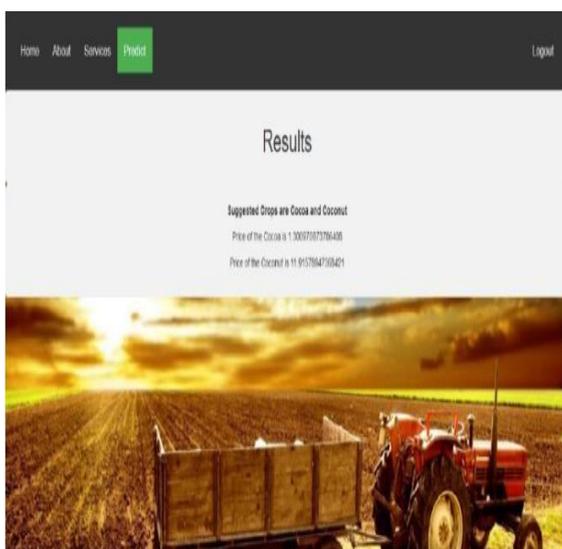


Fig. 5-: Result page

The Fig. 5 shows the Result page. The user inputs the value to the prediction then the result page will be displayed which will display the suggested crops and price of the suggested crops.

8. CONCLUSION

The work demonstrated the potential use of data mining techniques in predicting the crop yield based on the input parameters average rainfall and area of field. The developed webpage is user friendly and the accuracy of predictions are above 90 percent. The districts selected in the study indicating higher accuracy of prediction. The user friendly web page developed for predicting crop yield can be used by any user by providing average rainfall and area of that place. The process was adopted for all the area to improve and authenticate the validity of yield prediction which are useful for the farmers for the prediction of a specific crop.

9. FUTURE ENHANCEMENTS

The future work aimed at the analysis of the entire set of data and will be devoted to suitable strategies for improving the efficiency of the proposed algorithm. Use of such kind of approach to forecasting is not restricted to agriculture alone. The clustering and regression is one of the capable tool in field of data mining which can be used in several different ways.

REFERENCES

- [1]Rajshekhar Borate., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering, April 2016.
- [2] D Ramesh, B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology (IJRET), Vol.4, 2015.
- [3] Leemans V, M F Destain, "A Real Time Grading Method of Apples Based on Features Extracted from Defects", J. Jood Eng., 2004.
- [4] Ramesh A. Medar and Vijay. S. Rajpurohit "A Survey of data mining techniques for crop yield prediction", IJARCSMS, Volume 2, Issue 9, September 2014.
- [5] Veenadhari, S., Bharat Misra, D Singh, "Data mining Techniques for Predicting Crop productivity-A review article" IJCST, International Journal of Computer Science and technology march 2011.

[6]P.Surya, Dr.I.Laurence Aroquiaraj, "Crop yield prediction in agriculture using data mining predictive analytic techniques"International Journal of Research and Analytical Reviews(IJRAR),2018.

[7]Devika, B.Ananthi, Internatinal research journal of engineering and technology(IRJET),2018.