The Effective Yield Of Paddy Crop In Sivaganga District – An Initiative For Smart Farming

N. Gnanasankaran, E. Ramaraj

Abstract: Every Indian knows that agriculture is the back bone of our country. It is true that developed, underdeveloped and developing countries get their food from agriculture. Farming in India is done using mundane methods. It is a bottom line that most of the farmers lack exact knowledge in gaining yield in paddy cultivation makes their agricultural yield very low. A massive division of farming and agricultural activities are based on forecasting, which at time backslides. Since we know the benefits of proper soil moisture and its quality, air quality and irrigation, growth of crops, such parameters cannot be neglected. The main objective of this research work is to propose an innovative idea of crop monitoring and smart farming using a big data analytical model and IoT. This idea tries to digitize farming and agricultural activities so that the farmers can check on the needs of crops and accurately foresee their growth. This research work is being carried out in the Sivaganga District, situated in the southern part of Tamil nadu. The preliminary reported in this paper deals with innovative applications in data mining, machine learning domain models for analyzing datasets, tools implemented in machine learning techniques. The data collected from various places, to identify the key parameters are elaborated. Different types of classifications of soil in Sivaganga district is carried out. Paddy cultivation methods, paddy varieties mostly used in Sivaganga district and yield parameters were analyzed and finally the morphological characteristics of paddy crop are discussed in this paper.

Index Terms: Agriculture, Smart farming, Data mining, Big data, Machine Learning, IoT.

1 INTRODUCTION

Rice is the vital food crop in India. Rice is the primary source of food in more than half of the worlds population. More than 90% rice is grown and obsessed in Asia. Rice is being planted in more than 154 million hectares land skirts annually or on 11% of worlds cultivated area and it is consumed by more than three billion Asians. Based on the above statistical survey on the needs of rice that total requirement cannot be supplied by traditional rice varieties. Hence it is a high priority need that the yield of rice has to be increased around India and other Asian countries and new improved rice varieties have to be identified and grown in an extensive manner. With this motive this research work has been evoked to increase the yield of the paddy crop and with a small initiative it is started in Sivaganga district in south Tamil nadu. By digitizing the farming methods using big data and various machine learning techniques one can accurately predict the yield expectation of paddy crop which is the major objective of this work.

2 PRELIMINARY STUDY

The initial study deals with innovative and effective tools used in data mining, classification of datasets, data prediction and machine learning models divulge for the analysis of datasets. Waikato Environment for Knowledge Analysis (WEKA) is an effective software tool for developing innovative applications in agriculture using data mining and big data. It is a far-reaching suite for applying data mining to large datasets. The mined information is represented as a model of well formed structure of datasets. Later the tool is implemented for new data prediction and classification. Further machine learning technique is induced in domain models for analyzing datasets. The main objective of using WEKA tool is to mine information for existing datasets and to perform research in data mining by developing new machine learning algorithm. WEKA system incorporates a set of data pre processing routines and manipulation of raw data and its transformation into a form of data mining feature selection tools that includes identifying irrelevant data and excluding it from the existing datasets. Classifiers and algorithms handle categorical and numerical learning tasks. The meta classifiers enhance the performance of the algorithms. Various machine learning techniques can be implemented with WEKA platform,

- ZERO – R is a technique that can be used for dataset pertaining with single rule.
- ONE – R is a simple rule based techniques used on single attribute which is useful in generating a baseline for classification performance. [Holte 1993 & Garner 1995].
- Naïve Bayes technique produces probabilistic rules for the items belonging to each of the possible class categories. [Langlay 1992].
- Bayesian Classifier is an another naïve approach through which that attributes are treated as though they are completely independent. They provide equality among the models.
- Decision Table contains same number of attributes as the original datasets and a new dataset is assigned a category by finding the line in the decision table that matches a non class value of the data item. [Wrapper method – John & Kohavi 1997].
- Intense Based Learning Scheme creates a model by simply sorting the dataset. It implements K nearest neighbor classifier, [Aha 1992] using cross validation.
- J – 48 is a technique used for practical machine learning. It generates a decision tree by producing a set of rules called decision list. It is an ordered set of rules.
- Sequential Minimal Optimization Algorithm is used to support or enhance Support Vector Machine (SVM). [Dumais 1998].
- WEKA's Numeric Prediction can be succeeded by using Linear Regression, M5 Prime and Decision Stamp methods. [Wang & Witten 1997].
- As the meta classifier can enhance the algorithm performance there are some prominent methods to construct ensemble classifier. They are Boosting and Bagging, Clustering method. Boosting and Bagging is used to identify vote on classification using weighted vote and Clustering is used to gain semantic understanding of the relationship between data instances.

These methods were implemented for generating rules for mushroom cultivation in Newzeland. All the above mentioned techniques can be implemented in paddy crop cultivation using data mining and the past and present real time dataset collected for smart farming.

3 DATA COLLECTION AND PARAMETER IDENTIFICATION

Agriculture is dependent on various important factors to achieve fruitful yield of crops. It also includes several risk factors such as timely forecasts of crop production, weather conditions, disease and pest control, harvest operations and moreover it needs policy decisions like storage, distribution, pricing, marketing and import – export. Many new advanced data mining techniques have to be established to create meaningful relationships for crop yield and its estimation. A farmer harvests not only crops but also growing amounts of data. Farmer's main objective in his life is his farming lands soil condition which results in the yield. Further extended to the type of crop he cultivates and supportive climatic conditions. The agricultural yield in India relies on agro climatic conditions, resource endowment technology level, techniques adopted infrastructure, social and economic conditions, fertilizer and pesticide agencies. Different crops are grown in different regions in India under different conditions. Fluctuation in crop productivity creates severe distress to the farmers. So, the estimation and productivity of different crops is a challenging major activity which comprises of revenue, agriculture, economics and statistical study to estimate the yield process. Thus, new forecasting methods have to be implemented for crop yields in India. Totally around 70,000 hectares of agricultural land is available around Sivaganga district. In this, 50,000 hectares of land comprises of black soil/ clay soil and remaining 20,000 hectares comprises red soil. To exactly identify the crop yield, four important datasets are mandatory. These are: Soil dataset, weather dataset, rainfall dataset and pest dataset. These datasets have to be collected throughout the paddy cultivation period (starting from the sowing of seeds till the time of harvest). In the beginning, meteorological rainfall and weather datasets have been collected from CTC-CECRI, Mandapam and State Ground Surface Water Resource Data Centre in Chennai. Later, Soil testing and Water testing has to done to know the types of soil and its moisture level and evaporation level, similarly water testing is carried out to identify the pH value and the type of water (Hard water or Soft water). Basically two major types of irrigation were carried out. It is either irrigated (providing water source whenever needed either through pump set, bore or tank) or rain fed (depending upon the rainfall). [Data collected from Dry Land Agricultural Research Centre]. To modernize these procedures and to measure the soil and water type and level, IoT plays a vital role in monitoring the entire cultivation process. Initially the type of sensor has to be identified. Several type of sensors are available in the market such as Pest sensor, water and field sensor, arable sensor and soil sensors to analyze soil moisture, soil content and soil depth. As soil data plays a major impact in the field of agriculture, data collection for soil from nine different fields around Sivaganga district has been initiated and the work is in progress. After soil testing process is completed, using the soil sensor the growth of the paddy crop at various stages will be monitored.

4 MATERIALS AND METHODS FOR SOIL CLASSIFICATION

Soil types must be same in all locations for predicting accurate yield results. The survey manual of Indian Government has classified soil into eight classes. They are from class 1 to class 8.

Class - 1 and 2: Good for cultivation. It may have very fewer limitations such as sloping land and slight erosion.

Class - 3: Still can be cultivated with severe limitations such as moderate slope, erosion and shallow root zone.

Class – 4: Severe limitations but still can be cultivated with good management practices.

Class – 5: Nearly level but unsuitable as it may be dry, rocky or most often very wet. Suitable for pasture, wildlife habitat or forecast prediction. Class 4 and 5 are mostly similar.

Class – 6: Unsuitable as it may have high erosion and sloping land.

Class – 7: Severe limiting properties such as steep, severely eroded, deep gullies and very course. Can be turned into pasture but grazing must be controlled. Used as forecast/ recreation.

Class – 8: Extreme limitations. Left in its natural state. Based on the above study of soil variety and its classification, basically four important paddy crops are cultivated in Sivaganga district,

Samba (140 -145 Days of cultivation period)
BPT 5204 (140 Days of cultivation period)
(Deluxe Variety)
TKM 13 (135 Days of cultivation period)
KO 51 (105 Days of cultivation period)

Based on the above four paddy varieties intensive parameter identification was carried out. The key parameters identified are, seed variety, soil variety and type, soil moisture, onset monsoon, rainfall, seed suitability, duration, morphological characteristics, fertilizer application, macro and micro climate, pest surveillance, maturity, irrigation and productivity.

5 MORPHOLOGICAL STUDY ON PADDY CROP

As rice is an important food crop in India and to proceed with the next level of extensive research work it is vital to know the morphological characteristics of paddy crop. Morphology is a branch of Biology that deals with the study of forms and structure of organisms and their features. As per the statistical survey mentioned earlier in this paper, based on the need and highly consumed crop of India the total requirement of this crop cannot be supplied to the entire nation by following traditional methods and rice varieties. Hence there is a need for improved rice varieties.

Agricultural scientists try to increase the yield potential of rice by framing standard strategies for increasing the yield potential of rice through conventional hybridization, selection procedures, ideotype breeding, heterosis breeding, wide hybridization and genetic engineering. These methods can ultimately improve the morphology of paddy crop. By the successful implementation of the above methods tremendous improvements can be observed in the plant height, plant leaves, number of tillers per plant, plant type, pinnacle weight and increase in numbers, leaf area index, ration of leaf areas to grains and harvest index. All the above parameters can be improved to convert an unhealthy paddy crop into a healthy canopy. Table 1 Shows the Morphological Improvement of paddy crop after implementing the above mentioned methods. Table 2 and 3 shows the improved features of the high yielded rice variety.

TABLE 1: *Morphological Improvement*

<table>
<tr><th>Big Pinnacles</th><th>Fewer Tillers</th><th>Plant Height</th><th colspan="2">Plant Leaves</th><th>Ideal length</th><th>Ideal width</th><th>Ideal length of first leaf under top leaf</th><th>Leaf angles</th></tr>
<tr><td rowspan="5">250 spike lits per pinnacle</td><td rowspan="5">3 – 4 productive tillers</td><td rowspan="5">Early crop – 55 - 60 cm Late Crop – 60 - 76 cm</td><td>Erect Leaf</td><td>Before appearance of pinacles</td><td rowspan="5">40-44 cm early crop 42-60 cm late crop</td><td rowspan="5">1.2 - 1.4 cm</td><td rowspan="5">39-44 cm 37-43 cm</td><td rowspan="5">Flag leaf – 5 degrees 2nd leaf – 10 degrees 3rd leaf – 20 degrees</td></tr>
<tr><td>V – Shaped Leaf</td><td>Leaf blade stiffer, no droopiness</td></tr>
<tr><td>Narrow Leaves</td><td>Effective leaf area index</td></tr>
<tr><td>Flag leaf</td><td>50 cm (19.7 inches)</td></tr>
<tr><td>2nd leaf</td><td>55 cm</td></tr>
</table>

TABLE 2: *Productive tiller number per plant*

Early crop	15
Late Crop	14-19
Less tillers	For flowering
More tillers	Less number of pinnacle baring tillers
Fewer tillers	3-4
Panicle length	13.3-29.6 cm
Panicle height	60 cm from ground
Low Panicle position	60-70 cm ripining stage
Harvest Index	<0.5

TABLE 3: *Morphological Model of Super high yielded rice*

Flag leaf	50 cm
2nd leaf	Larger than panicle and over the top
3rd leaf	Equal to middle position of panicle
Plant height	100 cm
Clum length	70 cm
Erect flag leaf angle	5, 10, 20 degrees till mature
Plant type	Moderate tillering and compact
Panicle weight	5g(Grain weight per panicle) 2.7 million panicles per hectare

6 AGRICULTURAL LAND SELECTION TO MONITOR THE DEVELOPMENT STAGES OF PADDY CROP

After identifying the above mentioned dataset types the past and present data were collected and various parameters were identified to carry out extensive research on yield estimation of paddy crop. Going forward to the next stage, the real time data has to be collected to fulfill the dataset needs. Totally nine farmers and nine different agricultural lands from different villages around Sivaganga district have been chosen. Each land measures a total area of 2.5 acres to calculate the exact amount of seeds sowed, the amount of fertilizers utilized during cultivation and finally to estimate the exact yield gained in each 2.5 acres area of land. In each field the following details have to be collected to monitor the cultivation stages of paddy crop: Farmers Name, Fathers Name, Place-Village-Panchayat-Taluk-District, Survey number, contact details, Soil type, Paddy Variety, types of fertilizers used at the different stages of cultivation, field size, Irrigation type. At present, in all the nine fields the soil has been collected and given for testing. The test results are awaited to monitor the next stages of cultivation.

7 CONCLUSIONS AND FUTURE WORK

This research work will be a great boon to the farmers to predict the exact yield estimation of rice in their agricultural fields by following the rules generated in this work. The future work comprises of various stages such as, land preparation, methods of raising seedlings, plant nutrient and effects of growth, fertilizer sources, insect pests of rice, pest prevention, pesticide safety and agro chemical use, pesticide calibration, rice diseases, weed control, harvesting, threshing, and drying and storage. All the above stages have to be monitored and therefore using the above discussed parameters the datasets will be pre processed, analyzed and algorithms have to be developed for calculating the effective yield of paddy crop.

8 ACKNOWLEDGEMENTS

This research work has been carried out with the financial support of RUSA phase 2.0 grant sanctioned vide Letter No: F 24-51/2014-U, Policy (TN-Multi Gen), Dept of Edn, Govt of India, Dated: 9.10.18.

9 REFERENCES

[1] Developing Innovative applications in agriculture using data mining. (2007, Newzeland).
[2] Role of Data mining in Agriculture. (Kolhapur 2013).
[3] Importance in crop Insurance, American journal of Agricultural Economics.
[4] Economic importance of Kolhapur district, (Dat and Dorfan 1991).
[5] Normality of crop yield distribution, (Just and Weinenger 1998).
[6] Data mining of agricultural yield adata: A comparison of regression models (George Rub 2012).
[7] Classification of agricultural land soils: A data mining approach (V. Ramesh and Ram 2013).

[8] Agriculture, Horticulture, Environment and land use management.
[9] Rainfall variability analysis and its impact on crop productivity (D.R. Mehta 2010).
[10] Generalized software tool for crop area estimation and yield forecast (Roberto Benedetti 2014).
[11] Risk in agriculture: A study of crop yield distribution and crop insurance (Narsi Reddy Gayam).
[12] USDA Soil taxonomy.
[13] TN Agriculture weather Network.
[14] Crop Pest DSS.