JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ASSESSMENT OF INFLUENTIAL FACTORS AND PROGNOSTICATION OF CROP YIELD

[1]Sumanth N J, [2]Punith L S, [3]Shivakumar R M, [4]Mandar K P, [5]Mrs.Shylaja N S

[1]UG Student, [2]UG Student, [3]UG Student, [4]UG Student, [5]Assistant Professor

[1]Department of Computer and Engineering,

[1]Government Engineering College, K R Pet, Mandya, Karnataka, India

Abstract: In a agro-based country like India Forming Plays an important role in the financial growth of the country. crop selection is an important matter during agriculture planning. which depends on different factors such as market price, rate of production, transportation facilities and government policies. Many studied are carried by researchers on the prognostication of yield rate of crop, prognostication of climate, classification of soil and crops types for planning in forming using different machine learning techniques or statistics methods. If there is a plan to plant multiple crop at the same time using bounded land resource, then the crop selection is a challenging task. This article suggest a method named Assessment of influential factors and Prognostication of Crop yield to resolve crop choosing issue, and increase the crop yield rate over season and to achieve the financial growth of the country. This suggested method may improves the net crop yield rate.

Key words **– Prognostication, Liner regression, Decision tree.**

1. INTRODUCTION

Achieving the maximum crop yield using minimum agriculture field is the objective of cultivation planning in an agrarian country. Previous analysis of problems related with yield rate of crop can helps in maximizing crops yield. selector of crops could be applicable to reduce the losses when unfavourable situations may occur and this selector could be applied to increase yield rate of crop when the favourable growing conditions exists. increasing productivity rate of crop is a fascinating field of research for agro-meteorologists which play a important role in countries economy. mainly two factors which impact crop yield rate: first one is hybridization technology used to enhance the grade of seeds using genetic development, and second one is crop choosing based on the favourable or unfavourable situation.

Many agricultural planning related researches are carried out, the main aim of those research is to get an accurate and efficient model for the prognostication of crop yield, soil and crop classification, weather prognostication, prognostication of crop disease, categorization of crops based on growing stages. Machine learning and statistical techniques were prototyped. This article develops an advanced method to increase crop yield over various season.

2. PROBLEM STATEMENT

Crop selection is a significant issue for the farmers during different seasons. various factors such as rate of production and market price and environmental conditions are the influencing factors for crop selection. While most of the farmer have lack of knowledge that what type of crop is most suitable for their soil type, temperature, humidity etc. Due to this problem farmers fail to achieve best yield from crops by growing non suitable crops.

3. EXISTING SYSTEM

Number of work related to Machine learning algorithms has been carried-out that are applicable in the field of agriculture. offering best quality and affordable cost for the end-user by maximizing the farm production is the biggest challenge in the farming sector. It is analyzed that the farm production of about 25% gets wasted, and end-user does not get those products. The existing prototype also suggest the alternatives to reduce wastage of farm products. the recent work of, S. Pavani et.al. demonstrated a prototype in which the yield of the crop is prognosticated using the K-Nearest Neighbour algorithm by making the groups. It is shown that K-Nearest Neighbour clustering provides best result than Support Vector Machine or regression algorithm.

- Nishant et. al. prognosticates the yield rate of crop using advanced regression techniques like Kernel Ridge algorithms for the particular year. The Stacking regression help to increase the algorithm accuracy.
- the proposed system provides crop suggestion based on the area or land type.
- There is no fertilizer recommendation system for specific crop, land type and climate condition.

4. PROPOSED SYSTEM

- A model has been proposed by us in this project that addresses the problems related with existing methods. The main objective of the proposed system is to advice the farmers to increase the yield rate of crop also to recommend the most economic crop for the specific region based on rainfall, soil ph value, humidity, temperature and climatic condition etc.
- The proposed model assists in the selection of crop based on environmental and economic conditions, and to increase the yield of crop that will frequently help to reach the demand of food supplies of the country. The proposed model prognosticates the yield of crop by analysing influential parameters like temperature, rainfall, climate, soil type, season, etc. The best fertilizers to use are also suggested by this proposed system.
- The soil type, temperature, humidity, Ph value of soil etc are provided by user inputs. the model prognosticates the yield rate for a particular crop according to the requirement. The proposed model also suggest the most economic crop and recommend the suitable fertilizers to use.
- obtaining a better variety of crops that can be grown over any season is the main objective behind this project. the difficulties faced by farmers in selecting a crop can be overcome by using the proposed system that also help to increase the crop yield rate.
- The method also provides type of fertilizer that can be used for the specific soil type, humidity, temperature and crop type, this will help us to grow crops easily and reduce the money wastage on spraying unnecessary fertilizers

5. MERITS

• our proposed model prognosticates the yield rate of crop for the data sets of the given factors. Adding the farming and Machine Learning will come up with enhancement of the farming field by maximizing the rate of yield and the resources involved are optimized. The previous year data sets are the key parameters in forecasting the present performance.

•The right fertilizers for the crops are suggested by proposed system.

•Maximizing the crop yield, selecting efficient parameters, real-time analysis of crops, getting better yield rate and making good decisions are included in the proposed system.

6. MODULES

6.1 Collection of data:

Agriculture related data set is collected and it is evaluated for the presence of any missing values. The dataset used is consistent and complete without any missing values. This data set is further analysed for thorough perception and data clarity and the dataset is loaded to the crop advice system for further proceedings

6.2 Pre-processing:

The Agriculture dataset contains types of soil. In the pre-processing task we analyse the dataset by viewing all instance for each kind of soil type separately for better understanding of the crop with its appropriate soil type and other parameters. This leads in selecting the right crop at the proper time and thus maximize the yield rate

6.3 Classification and grouping

In classification and grouping step the classification process is done based on attributes namely soil type, humidity, temperature, ph value etc. Finally the grouping process is done for the crop attribute and classify what crop has to be grown based on the classification algorithms like Random forest, SVM and XGboost with their accuracy.

7. ALGORITHMS

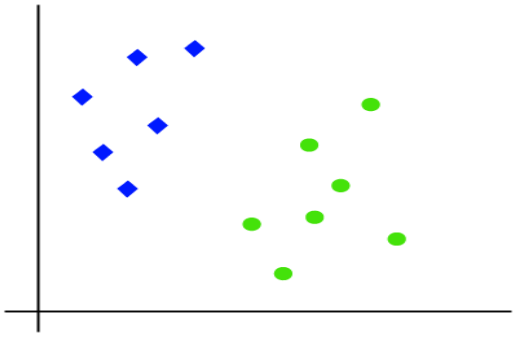
The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study,Data and Sources of Data, study's variables and analytical framework. The details are as follows;

we use different machine learning algorithms according to the dataset because the SVM Machine learning algorithm involves prognostication and classifying data.

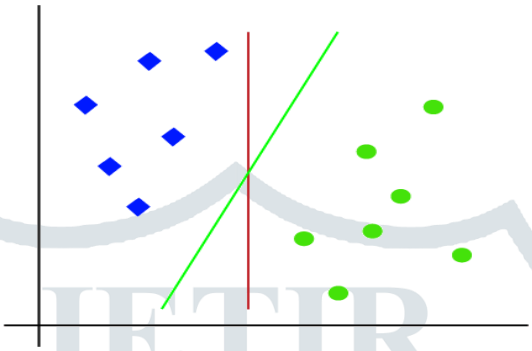
SVM algorithm is a linear model for regression and classification problems. this model work well for many practical problems and solve non-linear and linear problems. the simple idea behind the Support Vector Machine is that: The data gets separated into various classes by hyper plane created by the algorithm.

Let us consider a scenario.in case we have below shown data sets and we need to classify the blue rectangles from the green elliptical shapes (that is positives from the negatives). we have to find a line which is idle that separate these datasets into two different classes.

To separate this data sets into green and blue category we have to find a hyperplane or idle line. It looks like a simple task, but as we noticed that there is no single line to separate these data.



Let us consider some candidates of probability and highlight it.



Here we have two lines, blue and green.in the above graph, according to you which line best separates the data?

If you choose the green line then your answer is right, because we are looking for that line itself. In this case the green line classifies better. still, we need some solid reason to fix this line.

The red line in the above image is bit near to the blue rectangles. Although it is not a generalized line, it classifies the present data sets and Getting more generalized separator is the main goal of the machine learning.

Finding best line using Support Vector Machine's ways
we can find the closest points to lines from two classes using SVM algorithm. These closest points are referred as support vectors. The distance between the support vector and the line is called as margins. Maximize tis margin is the main goal. The optimal hyperplane is that hyperplane with maximum margin.

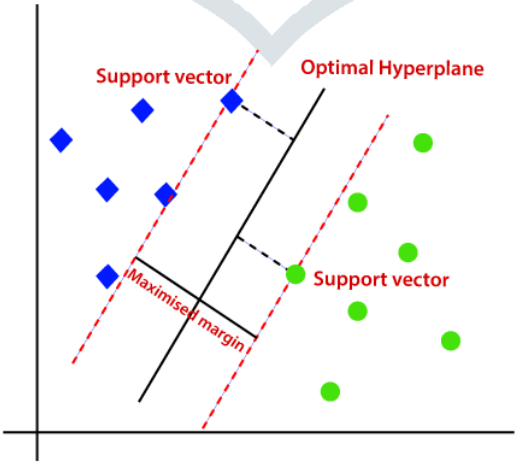
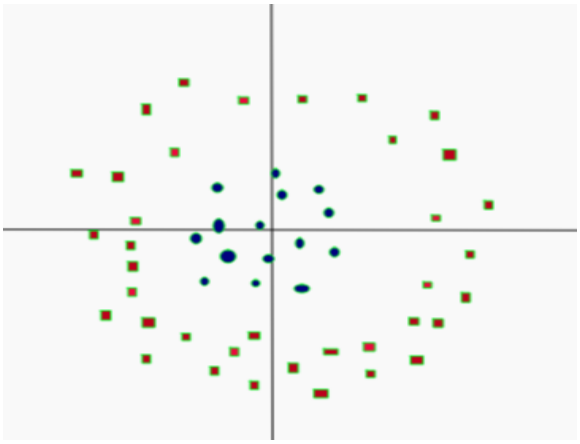


Fig 7.1 : Hyper plane

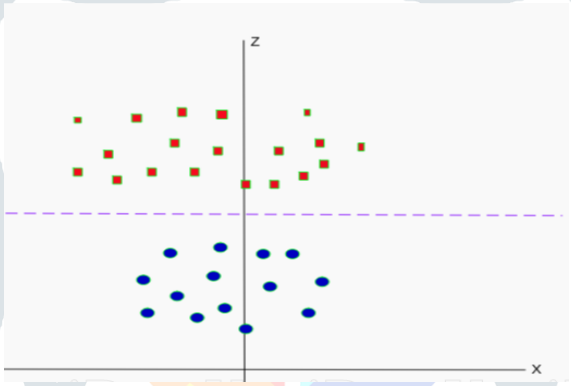
The support vector machine attempts to create a decision boundary in a way such that two classes separated as widely as possible. Let us take a quite complicated data set, that cannot be separated linearly



we cannot draw a straight line which can classify the data because we can saw that this data is non linearly separable data. Although in the higher dimension we can convert this data into linearly separable data. Now let us include another dimension that is z-axis. The constraints governs the co-ordinates on z-axis,

Z= $x^2 + y^2$

The z co-ordinate is basically sum of the squared distance to the points from the axis,



Now we can say that data is clearly linearly separable. We can write purple line as z = k which separates the data in the higher dimension, here k is the constant term. Let us write $x^2 + y^2 = k$, we know that z= $x^2 + y^2$; that is the equation of the circle. Using the transformation we can draw this linear separator into linear separator which is in the higher dimension.

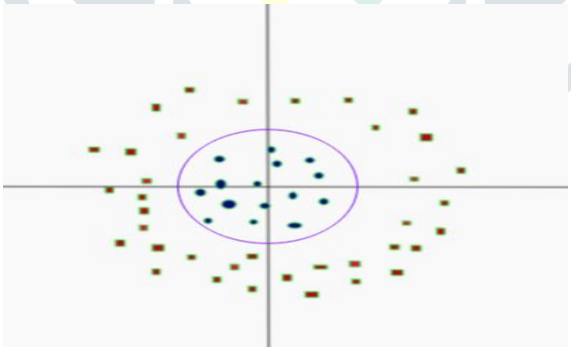


Fig 7.2: Original boundary containing the Decision Boundary

Using the mathematical transformation, we can project the decision boundary into original dimension and we can make data as linearly separable by adding extra dimension and classify the data. But it is very tedious task to find the correct transformation of any given data set. Surprisingly we have kernels in Sk-learn SVM implementation to complete this task.

Hyperplane

A hyperplane can be defined decision boundary that differentiates the two classes in SVM.

For the better understanding consider an example, in the Euclidian space we assume one-dimension line. Now make point on that one-dimension line so that point divides the entire line into two equal parts, here the point has 0 dimension and the line has 1 dimension so the point becomes the hyper plane of that one-dimension line.

Random Forest algorithm

Random forest algorithm is a supervised machine learning algorithm that is used in the regression and classification problems. On the different samples decision trees can be built in the random forest which uses majority of polling result for average in case of regression and classification.

One the most important feature of random forest algorithm is Handling the data set containing continuous variables in case of regression and Categorized variable in classification, better results can be calculated for the classification problems.

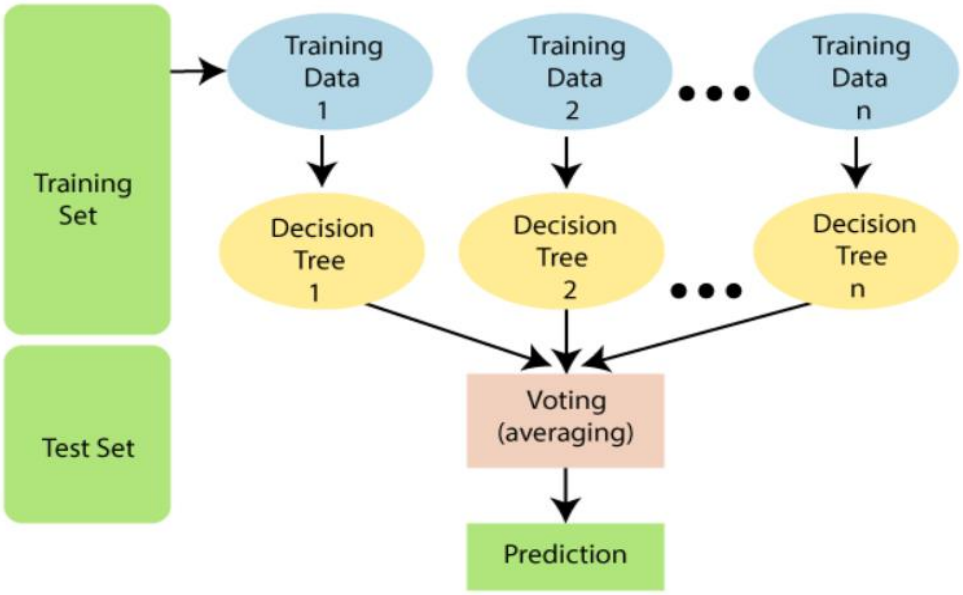


Fig 7.3: General Decision Tree

The random forest combines multiple trees to predict the class of data set

Working of random forest includes the following steps

Step-1: form the data sets containing K number of record, Random Forest takes N number of random records.

Step-2: For each sample individual decision trees are constructed.

Step-3: Each decision tree produce out puts

Step-4: The final out puts are used for classification and regression which is based on polling/majority voting or Averaging.

For example: consider a fruit basket as the data as shown in the figure below. Now N number of samples are taken from the fruit basket and for each sample an individual decision tree is constructed for. Each decision tree will generate an output as shown in the figure.

The final output is considered based on majority voting. In the below figure we can see that the majority of the voting in decision tree gives output as an apple of class A when compared to a banana of class B, so the final output is taken as an apple.

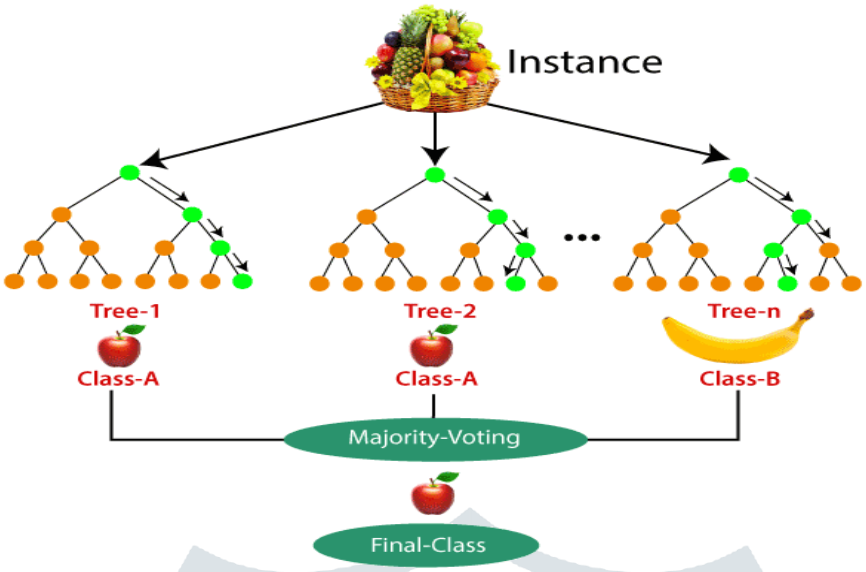


Fig 7.4: Fruit basket Decision Tree

Important Features of Random Forest

1. **Training, testing, splitting:** In the Random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
2. **Diversity:** during the construction of individual tress, we don't consider all the features or attributes because each tree is different.
3. **Stability:** as the results are based on the majority voting or polling, stability play as important role.
4. **Parallelization:** we can make use of CPU to create Random Forest because each tree is independently using different attributes and data.
5. **Resistent to course of dimensionality:** the feature space is reduced because all the trees doesn't consider all the features.
6. **Immune to course of dimensionality-** Since each tree does not consider all the features, the feature space is reduced.

8. CONCLUSION

The prognostication of yield rate of crop on the basis of NPK values and algorithm implementation has achieved that we can achieve better higher crop yield. From above work we conclude that Random Forest is good for soil classification.

For crop yield prognostication support vector machine is good with accuracy rate 99.47% compare to Random Forest. To implement the following functionality we can extends this work further. Different mobile applications can be built to suggest formers by uploading images of agriculture field. Crop disease detection uses processing of image so that farmer get pesticides based on image that reflect the disease of crop.

9. REFERENCES

[1]. Crop yield prediction using machine learning: A Systematic literature review ; Thomas van klompenburg ; ayalew kassahun ; cagatay catal , 2020.

[2]. Crop yield prediction using machine learning algorithms; Anakha venugopal; Aparna S; Rima Mathew; 2021.

[3]. Crop yield prediction using deep neural nural network; Saheed Khaki; lizhi wang; 2019.

[4]. A hybrid approach for crop yield prediction using machine learning and deep learning; Sonal Agharwal and Sandya Tarar; 2021.

[5]. Performance evalution of machine learning algorithm for mustard crop yield prediction from soil analysis; vaishali pandith ; Surjeeth sing; vinod Sharma; 2020.

[6]. Predicting the yield of crop using machine learning algorithm; P Priya ; U Muthaiah ; M Balamurugan; 2018.

[7]. Crop yield prediction using machine learning: A systematic literature review; Thomas van klompenburg; Ayalew Khassahun;2020.

[8]. Crop yield prediction using liner support vector machine;N. Manjunathan; P. Rajesh;A. Suresh , 2020.

[9]. Rice field forecasting Using support Vector Machine; sunil kumar; Vivek kumar; R.K Sharma;2019.

[10]. Efficient crop yield prediction using machine learning algorithm; arun kumar; Naveen Kumar; Vishal Vats; 2018.