Categorize the Quality of Cotton Seeds Based on the Different Germination of the Cotton Using Machine Knowledge Approach

P. Revathi [1], Dr.M.Hemalatha [2]

[1]Research Scholar, Karpagam University, Coimbatore, Tamil Nadu, India

.[2] Head Dept of Software Systems, Karpagam University, Coimbatore, Tamil Nadu, India.

kvrevathi1@gmail.com, hema.bioinf@gmail.com

Abstract

A study was made in those region that expose major area of machine learning techniques which available in data mining. There are lots of Research work is analysized in agriculture field. It is a smart techniques is needed to find the resolution over the modern method. Cotton is an significant beneficial harvest and extensively traded commodity across the world. Its yield is insightful to weather, soil as well as organization practices. This study presents a concise idea of several of the broadly capitalized machine learning technique over agricultural domain. Among the algorithms like Navie bayes, j48 ,MLP,Randomforest,Random tree, I have founded J48 gives better result over the other algorithms using cotton seed quality accuracy is best.

Keywords: *Machine learning, agriculture, Navie bayes, Random forest.*

1. Introduction

Machine learning research is learn to recognize tedious patterns automatically and make intelligent decisions based on data. The difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too huge to be covered by the set of observed training the seed data sets, and find the accuracy of seed quality capitalized from the java classifiers. Machine learning algorithms serve for inducing classification rules from a dataset of instances and thus augmentation the vicinity knowledge and perceptive.

Data mining software applications using various methodologies have been developed by both commercial and research centres [1]. These techniques have been used for industrial, commercial and scientific purposes. For example, data mining has been used to analyze large data sets and establish useful classification and patterns in the data sets. “Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and other analysis methods” (Cunningham and Holmes, 1999) [2].

The research aimed to establish if data mining techniques can be used to evaluate different classification methods by determining whether meaningful patterns exist across different of seed germination and seedling growth along with root zone temperature gradient different research sites. The data set has been assembled from cotton seed growth on stages at agriculture areas located in Coimbatore. The *researcher* has utilized existing data collected from agricultural field. The analysis of these agricultural data sets with different data mining techniques may yield outcomes useful to researchers in the seed science. It is envisaged that the information gained from these research will contribute to the improvement and maintenance of seeds and the agricultural environment of seed science.

In this work described the remarkable of machine learning classifier in agriculture database and extracting knowledge, and compared these classifiers and finally got best solution.

2. Development of Cotton Seedling

Cotton germination begins as the seed absorbs water and oxygen through its chalaza after planting. The water swells the dormant tissues, and cell growth and division begin to take place. The radicle emerges through the micropyle, turns downward, and grows deeper into the soil, providing a taproot that will supply water and nutrients throughout the life of the plant. The hypocotyl elongates from the radicle and forms an arch or crook that begins to push up through the soil, a brief period often referred to as the “crook stage” [3].

Cotton seedling emergence normally takes place 4 to 14 days after planting. At the soil surface, the hypocotyl straightens and pulls the folded cotyledons out of the soil, a process known as epigeal germination. After the cotyledons are pulled through the soil surface, they unfold and expose the epicotyl and the apical meristem, or growing point, which will be the source of subsequent growth. At this point, germination and seedling emergence are complete and the plant begins its active.

2.1. Development of Roots

Root growth dominates the growth of the cotton plant during germination and seedling establishment. In fact, the taproot may be as deep as 10 inches by the time the cotyledons emerge. This is a critical time for the development of the root system. Cold soils, seedling disease, low soil pH, water stress, hard pans and herbicide injury all inhibit root growth and development, but careful crop management can minimize most of these stresses. The roots absorb water and nutrients that are vital to the development of the plant and any hindrance of root development in these early stages of cotton growth may cause a disappointing production season [3].

Cotton germination the quickest from warm, moist soil. Low temperatures (below 60 degrees F) or less than adequate soil moisture may hinder germination by slowing metabolic processes (see the discussion on heat units). Physical impedance, such as crusting, does not slow germination, but it can prevent the hypocotyl from emerging. This often causes thickening of the hypocotyl and a condition referred to as “big shank” or “thick-legged” cotton, resulting in reduced seedling vigor. Generally, the longer it takes for emergence to occur, the greater the risk of plant death and yield loss. A rule of thumb for planting cotton in most regions of the U.S. Cotton Belt is that the soil temperature at 4 inches deep should be at least 65 degrees F for 3 consecutive days, with warm temperatures in in the forecast.

3. Motivation

Cotton needs sustainable cultivation practices. The unpredictability and delay in the present production estimations are posing serious problems to planners to take timely import–export decisions. consistent prediction methods are therefore needed to help planners and policy makers take strategic decisions to safeguard national interest. Transfer of technology is one of the important areas to be addressed. Most of the work so far carried out concentrates on different dimensions. A methodology is described in [4] to predict cotton production on a regional basis using the integrated approach of remote sensing (RS), geographic information system (GIS) and a crop simulation model, i.e. Infocropcotton model. The study in [5] monitored variations of seed germination and seedling growth of four important crop plants namely Cotton, Wheat, Rice and Maize in Egypt along root zone temperature gradient. The management of cotton yield behavior inagricultural areas is a very important task because it influences and specifies the cotton yield production. An efficient knowledge-based approach utilizing the method of fuzzy cognitive maps (FCMs) for characterizing cotton yield behavior is presented in [6].A very few work has been carried out using the promising field of machine learning.

4. Machine Learning Classifiers

Machine knowledge is concerned with creating and using mathematical “data structures” that concur to a computer to expose performance that would generally require a human. The subsequent machine knowledge classifiers are used in classification problem to construct the model using training data and to categorize the quality of cotton seeds [7]. The data structures used in classification problems, to build models come in various forms, eg. trees, Graphs, probability distributions. The following machine learning classifiers are used in classification problem to build the model using training data and to classify the cotton seed based on the germination stages of cotton crop.

1. Naïve Bayes:

The naive bayes techniques is extremely admired approach even if it is (apparently) based on an unrealistic assumption.The distributions of the predictors are mutually sovereign conditionally to the values of the goal attribute. The main motive of this popularity is that the procedure proved to be as exact as the other well-known approaches such as linear discriminant analysis or logistic regression on the majority of the real datasets. The Naive Bayes classifier (NB) is a simple but efficient classifier, which has been used in numerous applications of information processing including, natural language processing, information retrieval, etc. The Naive-Bayes inducer computes conditional probabilities of the classes given the instance and picks the class with the highest posterior [8]. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting.

2. Decision Tree:

Decision tree is the most important representative of the family of symbolic machine learning techniques. A decision tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes from a given data set. Decision tree learning is the process of learning decision trees from the labeled training examples. Decision tree classification algorithm generates the output as a binary tree like structure called a decision tree, where each non leaf node i.e., internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node or terminal node holds a class label. The topmost node in a tree is the root node. A decision tree model contains rules which are used to predict the target variable. The class label of a new instance is predicted by testing the attribute values of the instance against the decision tree. A path is traversed from the root to a leaf node, which gives the class label of that data. Decision trees can be easily converted into classification rules.

3. Multilayer Perceptron:

A multilayer perceptron is principle weakness of the perceptron was that it could only solve problems that were linearly separable and feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. The multilayer perceptron consists of an input and an output layer with one or more *hidden layers* of nonlinearly-activating nodes. Each node in one layer connects with a certain weight *wij* to every node in the following layer.

5. Experiment and Results

Classification of cotton seeds on the Germination stages of Cotton crop has been passed out using implemented in java tool [9].The information necessary for the classification namely, family, varieties, growing zone, germination days, squaring days, flowering days, bursting days, soil type, weather (temperature and solar radiation), crop management variables (soil fertility, pest control, weed management) and so on were gathered. From the raw data set, 24 distinctive features were extracted for each instance, and stored as a fixed length vector for cotton seed yield analysis. The features mainly include culture, germination days, squaring days, flowering days, boll maturity days, bursting days, total days, family, zone, soil properties, dates of sowing, growth days, maturity, harvesting, irrigation, pest management, boll worms measurement, levels of fertilizers.

The cotton seed dataset consists of 500 instances, with 24 attributes of varying quality - Good, Average and Bad. The features are extracted from the cotton seed growth dataset and are used for constructing the appropriate model. The training set consists of equal number of instances of cotton seed of dissimilar quality. As the quality of the cotton seed has to be predicted, Quality- the categorical attribute is selected as the class label. The instances in the dataset pertaining to the three quality categories Good, Average and Bad are designated as Good, Avg, Bad to represent the quality level of the seed.

Machine learning scenario using java implements tool and compared these classifier after that owing the best solution of decision tree from other classifiers. The results of the experiments are summarized in Table 1.and performances of these models were evaluated based on predict the accuracy , learning the time build the model and error rate and illustrated in figures 1, 2 and 3.

Table1. The Comparative Result of the Classifiers.

Evaluation Criteria	**Naïve Bayes**	**J48**	**MLP**	**Random Forest**	**RFF Network**	**Random Tree**
Timing to build model (in secs)	0.02 Sec	0 Sec	0.27 Sec	0.06 Sec	0.27 Sec	0 Sec
Correctly classified Instances	490	499	488	478	462	451
Incorrectly Classified Instances	9	0	11	21	37	48
Prediction Accuracy (%)	98.19%	100%	98.78%	95.79%	92.58%	90.38%

Figures.1. Shows that J48 decision tree predicts the best performance from other classifiers used for experiments.

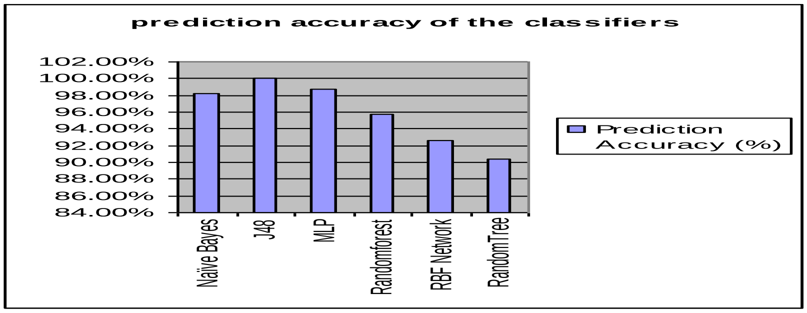

Figure1. Predict Accuracy of the Classifiers.

Figure.2. Represent the error rates and correctly classified instances, incorrectly classified instances of the data sets these classifier. J48 classifier is correctly classified instance. So decision tree owing the best performance among the other classifiers. Can be employed in the agriculture databank to predict the quality of the cotton seed that will enable increase the cotton productivity and aid in cotton seed certification.

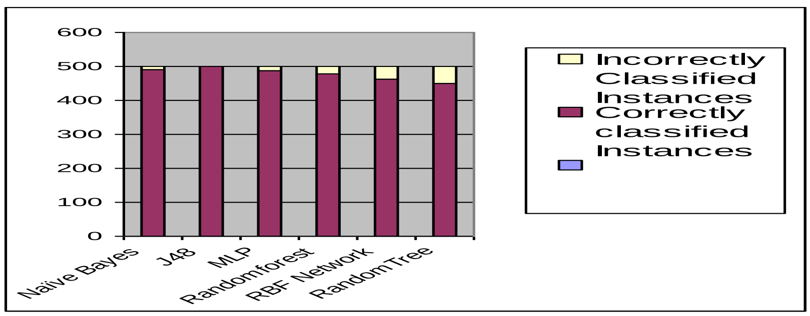

Figure2. Error Rate of the Classifiers.

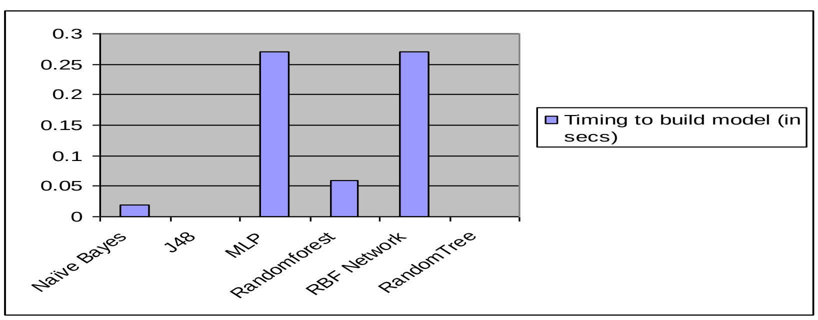

Figure3. Time Taken Build the Models.

Figure.3.Represent the time taken to build the J48 and Random Tree very less secs from other classifiers.

6. Conclusion

Machine learning is a burgeoining new expertise with a wide range of prospective application, This paper represent a cotton seed accuracy found the decision trees in discovering classification rules from the data of the cotton seed as the key crop in the agriculture field particularly, the huge numbers of expert system that have been developed for agricultural problems worldwide present further evidence that formalizing knowledge can benefit agriculture. It seems likely machine learning can contribute to the economic on several different fronts. Widely used supervised machine learning techniques namely J48 decision tree classifier, multilayer perceptron, navie bayes, Random Forest, RFF Network, Rand Tree classifiers were used for learning build model, features extracted from a set of agriculture record sets of different categories were used for training and implementation. The results of the models were compared these classifiers,J48 was better accuracy from other algorithms.

References

[1] A survey of Knowledge Discovery and Data Mining process models The Knowledge Engineering Review, Vol. 21:1-2006, Cambridge University Press Printed in the United Kingdom.

[2] Cunningham, S. J., and Holmes, G. (1999). *Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999.*

[3] http://landofcotton.com/fc/files/products.pdf

[4] K. B. Hebbar1, et.al, "Predicting cotton production using Infocropcotton simulation model, remote sensing and spatial agro-climatic data", Current Science, Vol. 95, No. 11, pp 1570 – 1579 , December 2008.

[5] Mahmoud A.Medany, et al.,"Prediction of seed germination and seedling growth of four crop plants as affected by root zone temperature", World Journal of Agricultural Sciences 3 (6): 714-720,2007.

[6] Elpiniki I.Papageorgiou, Athanasios Markinos, Theofanis Gemptos, "Application of fuzzy cognitive maps for cotton yield management in precision farming",Expert systems with Applications, Volume 36, Issue10, pp 12399 – 12413, 2009.

[7] Ian H. Witten and Eibe Frank (2005), " Data Mining: Practical Machine Learning Tools And Techniques" 2nd Edition. San Fransisco: Morgan Kaufmann.

[8] Lan H. Witten, Eibe Frank,"Data Mining - Practical Mahine Learning Tools and Techniques," 2nd Edition,Elsevier,2005.

[9] Ian H. Witten, Eibe Frank, et al., "Weka: Practical Machine Learning Tools and Techniques with Java Implementations", Working Paper 99/11, Department of Computer Science, The University of Waikato, Hamilton, 1999.

Authors

Dr. M. Hemalatha completed MCA MPhil., PhD in Computer Science and currently working as a Asst Professor and Head, dept of software systems in Karpagam University. Ten years of Experience in teaching and published Twenty seven papers in International Journals and also presented seventy papers in various National conferences and one international conferences Area of research is Data mining, Software Engineering, bioinformatics, Neural Network. Also reviewer in several National and International journals.

P.Revathi was born an brought up in coimbatore.she received her Bachelor of Business Administration from Madras University. and Master degree in Computer Applications from Bharathiyar University. She completed her M.Phil from P, S, G, R, K rishnammal colleges for women. Currently pursuing Ph.D., in computer science at Karpagam University under the guidance of Dr.M.Hemalatha Head, Dept of Software System, Karpagam University, Coimbatore. Area of Research is Data mining and Machine learning.