International Journal of Agriculture Innovations and Research
Volume 1, Issue 4, ISSN (Online) 2319-1473

A Web-Based Soya Bean Expert System Using Bagging Algorithm with C4.5 Decision trees

Prof. M.S. Prasad Babu
Dept. of CS & SE,
Andhra University,
Visakhapatnam, A.P, India.
msprasadbabu@yahoo.co.in

Swetha Reddy
M.Tech, Dept of CS & SE,
Andhra University,
Visakhapatnam, A.P, India.
swetha04552@gmail.com

B. Venkata Ramana
Assoc Prof., Dept. of IT,
AITHAM, Tekkali,
Srikakulam, A.P, India.
bendi.ramana@gmail.com

N. V. Ramana Murty
Assoc. Prof., Dept of MCA,
G.V.P.College (A),
Visakhapatnam
mursat78@gmail.com

***Abstract* – The Machine learning [1] is a mechanism concerned with a computer program that automatically improves its learning capabilities with experience. The Bootstrap aggregation (Bagging) is one of the most popular ensemble methods in Machine Learning. Bagging algorithm uses any classification method to increase the performance of the classifier. In this paper, bagging algorithm is used to increase the performance of the C4.5 classifier. Bagging algorithm runs the slightly altered data on a given classification method several times and combines the hypotheses for achieving higher accuracy in the simple classification method. A knowledge Base known as 'Soya Bean Expert Knowledge Base' is constructed by conducting programmed interviews with domain experts in Soya Bean crop production. A Soya bean expert system is developed to identify the disease of the crop with the use of bagging algorithm. A separate user interface for the Soya Bean expert system, consisting of three different interfaces namely, End-user/farmer, Expert and Admin is presented here. End-user/farmer module may be used for identifying the diseases for the symptoms entered by the farmer. Expert module may be used for adding rules and questions to data set by any domain expert. Admin module may be used for maintenance of the system. This expert system is a web based application for online users with JSP as front end and MYSQL as backend.**

***Keywords* – Machine Learning, Expert System, Rule Based System, Bagging Algorithm, Decision Trees, C4.5, Soya Bean, JSP, MySql.**

I. Introduction

A. Machine Learning

Machine learning is concerned with the design and development of algorithms that allow computers to evolve intelligent behavior based on empirical data obtained from sensors or databases. The key issue in the development of Expert Systems is the knowledge acquisition for building its knowledge base. One simple technique for acquiring the knowledge is direct injection method in which the knowledge is collected from the domain experts by conducting programmed interviews and entering it in an appropriate place manually. But it is difficult process and time consuming. Instead, machine learning algorithms are used by making the systems learn from their past experiences. The goal of machine learning is to program computers to use training data or past experience to solve a given problem. Effective algorithms have been invented for certain types of learning tasks. Many practical computer programs have been developed to exhibit useful types of learning and significant applications have begun to appear. Machine learning refers to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. Some of the machines learning algorithms are Genetic Algorithm, ID3, ABC algorithm, Artificial Neural Networks and C4.5 Algorithm etc.

B. Expert Systems

An expert system is a computer system that emulates the decision-making ability of a human expert. Expert systems have emerged from early work in problem solving, mainly because of the importance of domain-specific knowledge. The expert knowledge must be obtained from specialists or other sources of expertise, such as texts, journal articles, and data bases. Expert system receives facts in the form of queries from the user and provides expertise in return. The user interacts with the system through a user interface, constructed by using menus, natural language or any other style of interaction. The rules collected from the domain experts are encoded in the form of Knowledge base. The inference engine may infer conclusions from the knowledge base and the facts supplied by the user. Expert systems may or may not have learning components. A series of Expert advisory systems [12], [13], [15] were developed in the field of agriculture and implemented in www.indiakisan.net[14].

C. BootstrapAggregation(Bagging)Algorithm

Bagging is a Meta algorithm developed by Breiman (1996). The name derived from "bootstrap aggregation", was the first effective method to improve machine leaning of classification and regression model in terms of stability and classification accuracy. The meta-algorithm, which is a special case of the model averaging, was originally designed for classification and is usually applied to decision tree models. The method uses multiple versions of a training set by using the bootstrap, i.e. sampling with replacement. Each of these data sets is used to train a different model. The outputs of the models are combined by averaging the data (in case of regression) or voting (in case of classification) to create a single output. Bagging is effective when using unstable nonlinear models (i.e. a small change in the training set can cause a significant change in the model). It also reduces variance and helps to avoid over fitting. Although it is usually applied to decision tree models, it can be used with any type of model. In this proposed system the Classifier is C4.5 decision tree algorithm.

The training set D of size n, bagging generates m new training sets D_i, each of size $n' \leq n$, by sampling examples from D uniformly and with replacement. By sampling with replacement, it is likely that some examples will be repeated in each D_i. This kind of sample is known as a bootstrap sample. The *m* models are fitted as single model by using the above *m* bootstrap samples and combined by voting.

Input: $(X_{i,} Y_i)$ i=n;

X_i→ instance vector

Y_i→ output to particular instance

The C4.5 takes (X_i, Y_i) as input it generates *m* classifiers (bootstrap samples).

```
•   Algorithm Bootstrap-Aggregation (D, L, k)
        FOR i ← 1 TO k DO
            •   S[i] ← Sample-With-Replacement (D, m)
            •   Train-Set[i] ← S[i]
            •   P[i] ← L[i].C4.5 (Train-Set[i])
        RETURN (Make-Predictor (P, k))
•   Function Make-Predictor (P, k)
        RETURN (fn x ⇒ Predict (P, k, x))
•   Function Predict (P, k, x)
        FOR i ← 1 TO k DO
            Vote[i] ← P[i](x)
        RETURN (argmax (Vote[i]))
•   Function Sample-With-Replacement (D, m)
        RETURN (m data points sampled i.e. uniformly from D)
```

D. C4.5 Decision Tree Algorithm

C4.5 is an algorithm, developed by Ross Quinlan, used to generate a decision tree. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. C4.5 builds decision trees from a set of training data in the same way as that of in ID3, using the concept of entropy and information gain. The attribute with the highest normalized information gain is chosen to make the decision.

C4.5 Algorithm Steps:

1. Check for base cases
2. For each attribute A
    1. Find the normalized information gain from splitting on A
3. Let A_best be the node that will have highest normalized information gain
4. Create a decision node that splits on A_best
5. Recurse on the sub lists obtained by splitting on A_best, and add those nodes as children of node.

II. Knowledge Base

Knowledge base of an expert system contains a formal representation of the information provided by the domain experts. The information is collected from the domain experts by conducting programmed interviews. The Knowledge Base of the Soya Bean Expert System contains the information about the diseases and the symptoms for corresponding diseases and the cure for those diseases. The symptoms and diseases occurred in Soya bean crop are represented in tabular format as below.

S No	Symptoms	Disease	Cure
1.	Reddish-brown lesions on the leaf and stem, reddish-brown lesions on the stem.	Diaporthe stem canker.	Stem canker can be reduced by delayed planting and foliar fungicides. Tillage may reduce disease problems in fields where this disease has been a problem.
2	Brown lesions may form on the hypocotyls of emerging seedlings. Black specks form inside the lower stem. Reddish-brown discolorations develop in the pith..	Charcoal rot.	Manage fields to reduce or avoid drought stress. Reduced tillage may reduce, perhaps due to cooler soils. Rotation with crops that have relatively low susceptibility to charcoal rot, may be beneficial. Reduced seeding rates may also reduce drought stress and charcoal rot
3.	Sunken lesions on stems and roots near the soil line, Seedlings or older plants	Rhizoctonia root	Encourage seedling health with good agronomic practices and the use of high quality seed.
4.	Seedlings can be attacked and killed in the ground	Phytophthora rot	Treatment of seed with the highest labeled rates of fungicidal compounds such as mefenoxam (Apron XL®) or metalaxyl (Apron®)
5.	Causes browning of the pith in the center of the stem, leaves may also develop brown and yellow	Brown stem rot	Use of resistant soybean varieties and rotation to non-host crops
6.	Leaves have white to light grey sports, powdery patches.	Powdery mildew	Foliar fungicides can help to manage this disease,
7.	White thin lesions along leaf surface and green tissue in plants	Downy mildew	Spray Dithane M-45 @ 2-2.5 gm/liter
8.	Bushy appearance due to proliferation of tillers which become chlorotic and reddish	Brown spot	Seed treatment with peat based formulation (Pseudomonas fluorescence) @ 16 g/kg of seed
9.	Irregular section of epidermis and Perforated Leaves.	Bacterial blight	Spray of Chelamin450 @ 2.5 to 4.0 g/liter of water
10	The affected area just above	Bacterial pustule	Spray of Sheethmar (Validamycin) @ 2.7 ml

	the soil line is brown water-soaked soft and collapsed		/liter water
11	Affected internodes become disintegrated and the presence of small pin-head like black sclerotic on the rind of the stalks	Purple seed stain	Spray of Zineb@ 2.4 to 4.0 g/liter of water (2-4 applications) at 8-10 days interval

The information provided in the above table is used to construct the following rules. These rules are stored in the knowledge base in the as a data base table using MySql database formats.

Rule1: S1=6, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Diaporthe stem canker**".

Rule2: S1=5, S2=0, S3=1, S4=1, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Charcoal rot**".

Rule3 S1=3, S2=1, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Rhizoctonia root**".

Rule4: S1=2, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Phytophthora rot**".

Rule5: S1=6, S2=1, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=1, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Brown stem rot**".

Rule6: S1=6, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Powdery mildew**".

Rule7 S1=0, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Downy mildew**".

Rule8: S1=4, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Brown spot**".

Rule9: S1=2, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Bacterial blight**".

Rule10: S1=5, S2=1, S3=0, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Bacterial pustule**".

Rule11: S1=6, S2=1, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=1, S12=0, S13=0, S14=1, S15=0, S16=1, S17=6, resultant disease may be "**Purple seed stain**".

III. Proposed Bagging Algorithm

The Proposed Bagging Algorithm (bootstrap aggregating) uses the C4.5 classifier as the Classification algorithm used to improve the classification accuracy. Briefly, the method works by training each copy of the algorithm on a bootstrap sample, i.e., a sample of size S chosen uniformly at random with replacement from the original training set T (of size S). The multiple classifiers that are computed from c4.5 classification algorithm are then combined using simple voting; that is, the final composite hypothesis classifies.

Flowchart for Bagging Algorithm

The Proposed Bagging algorithm uses C4.5 classifier and it uses the training data, multiple times the classifier (C4.5) is called based on the given input and the weights for each classifier and it generates a new hypothesis. Based on the final hypothesis, the algorithm classifies the given input into corresponding disease.

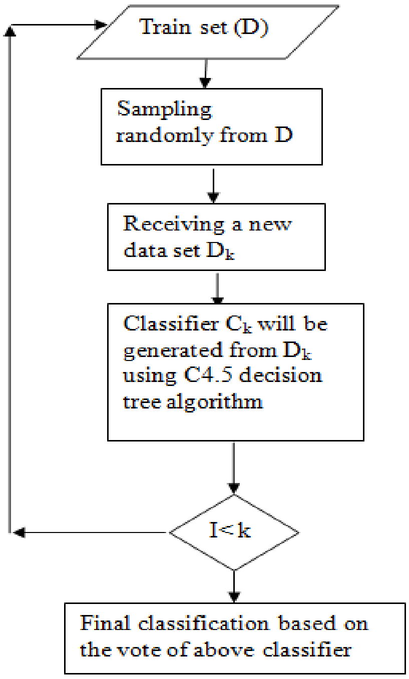

Train set (D)
Sampling randomly from D
Receiving a new data set D_k
Classifier C_k will be generated from D_k using C4.5 decision tree algorithm
I<k
Final classification based on the vote of above classifier

Fig.1. Flow Chart for Bagging Algorithm

A. Implementation of Rule Based System

Step 1: Rule based algorithm: Pseudo code for the implementation

Step 1.1: Enter the Symptoms to obtain the major disease.

Let us take two examples

Ex1 6,0,1,0,1,1,1,0,0,1,1,0,0,1,0,0,0

From rule 1 s1=6, s3=1, s10=1 and s15=0

Step 1.2: If the exact match symptoms found in the Knowledge base (KB) then Rule based system produce the output

=disease name as D1(Diaporthe-stem-canker)

Ex2 4,1,1,0,0,1,0,1,0,1,0,0,0,1,0,0,6

Step 1.3: Exact match symptoms were not there in KB

= Insufficient knowledge it fails and goes to System 2.

B. Implementation of bagging method

Step.2. Bagging Algorithm:

Step 2 1: Initialize the TEST SET

= Supply symptoms to obtain the major disease

Step 2.2: By using C4.5 decision tree classifier the training data is classified

It creates N decision trees with different attribute as root node.

= Classification based on the values of attributes in training data.

Step 2.3: Final classification is based on the majority vote of above N classifiers.

= classification model obtained on whole data set.

The disease which is having highest accuracy is taken and it is submitted to the user by the system as the disease affected to the crop.

Running Bagging Algorithm on given data it may generate a new rule, in that case that new rule is added to the list of rules in the database table. These rules may use full in future for better accurate results.

The following are the new rules generated so far by the Bagging algorithm.

Rule1: S1=4, S2=1, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1, S11=0, S12=0, S13=0, S14=1, S15=0, S16=0, S17=6, resultant disease may be “**Diaporthe stem canker**”.

Rule2: S1=5, S2=1, S3=0, S4=1, S5=0, S6=1, S7=1, S8=1, S9=0, S10=1, S11=1, S12=0, S13=1, S14=1, S15=0, S16=0, S17=6, resultant disease may be “**Charcoal rot**”.

IV. Comparative Study

In this section we compare the classification accuracy results of the two algorithms namely bagging with C4.5 Decision trees and C4.5.

Table 1: Performance of bagging

No of Iterations	Bagging with C4.5(accuracy)	C4.5(accuracy)
2	82.75%	77.69%
3	83.06%	78.06%
4	85.77%	78.77%
5	86.75%	80.75%
6	86.75%	80.75%
7	87.80%	81.90%
8	88.38%	83.38%
9	88.94%	84.94%

Table I demonstrate the classification accuracy results of two classification algorithms. It is evident from the table I that Bagging has the highest classification accuracy (88.94%). The classification accuracy for C4.5 algorithm is 84.94%. So the Bagging with C4.5 gives the better results than simple C4.5.

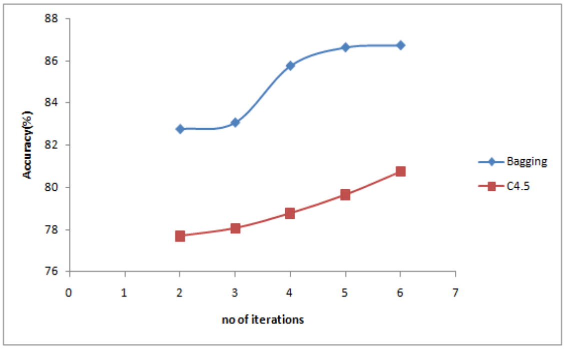


Fig.2. Comparison of performance

By using the above graph, it is clear that the performance of Bagging with c4.5 Classifier is better than C4.5 algorithm

V. Soya Bean Expert System Architectre

The Proposed architecture of the Soya bean Expert system consists of Rule based system, Bagging algorithm and knowledge base. It is represented in the figure 3

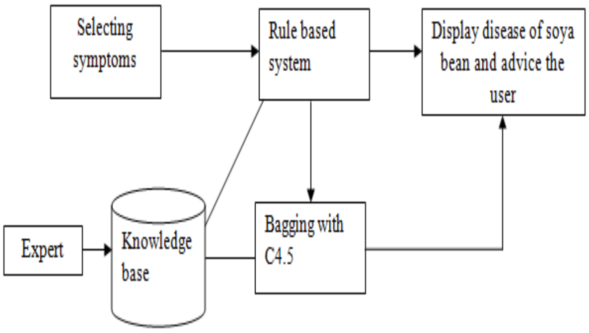
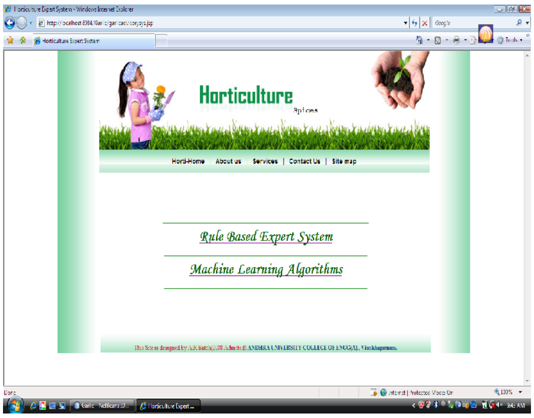


Fig.3. Proposed architecture of Expert System

VI. Results and Discussion

Test Results:


Horticulture Expert System - Windows Internet Explorer
http://localhost:8084/Garlic/garlicadvisorysys.jsp
Google
Horticulture Expert System
Tools
Horticulture
Spices
Horti-Home
About us
Services
Contact Us
Site map
Rule Based Expert System
Machine Learning Algorithms
Done
Internet | Protected Mode: On
100%
Garlic - NetBeans ID...
Horticulture Expert ...

This Site is designed by AIR Batch(2008 Admitted) ANDHRA UNIVERSITY COLLEGE OF ENGG(A), Visakhapatnam.

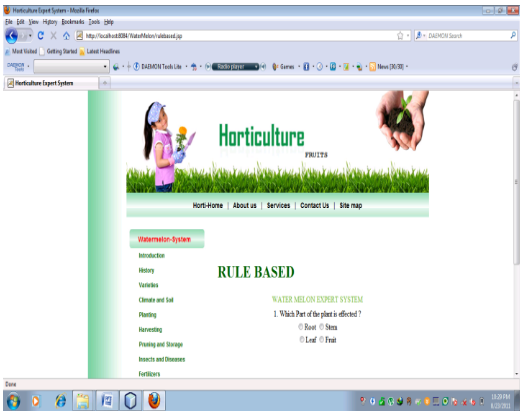
Fig.4. Selection of System

Description:

The user selects one of the two expert systems:

Rule based system
Machine learning system

Rule based expert system for Soya Bean



RULE BASED

WATER MELON EXPERT SYSTEM

1. Which Part of the plant is effected ?
○ Root ○ Stem
○ Leaf ○ Fruit

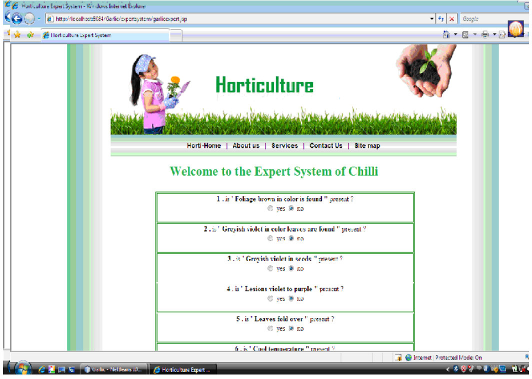
Fig.5. Rule Based System

Description: In this screen shot, the user can submit the observed symptoms to the Soya Bean rule based system through online by selecting the appropriate radio buttons for the processing of the symptoms observed.

Effected With: **Brown spot.**

Cure is: Seed treatment with peat based formulation (Pseudomonas fluorescence) @ 16 g/kg of seed.

Machine learning Expert System



Welcome to the Expert System of Chilli

1 . is ' Foliage brown in color is found '' present ?
○ yes ● no

2 . is ' Greyish violet in color leaves are found '' present ?
○ yes ● no

3 . is ' Greyish violet in seeds '' present ?
○ yes ● no

4 . is ' Lesions violet to purple '' present ?
○ yes ● no

5 . is ' Leaves fold over '' present ?
○ yes ● no

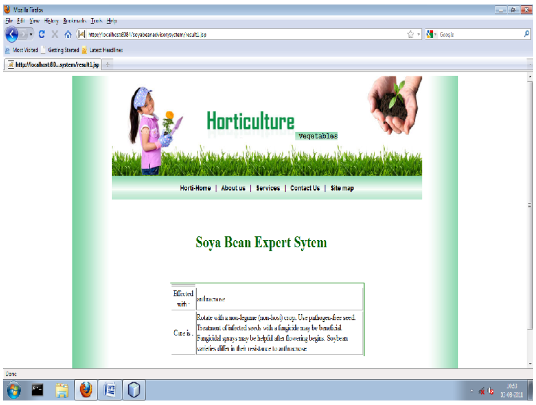
6 . is ' Cool temperature '' present ?

Fig.6. Machine Learning Expert System

Description:

In this screenshot, the user selects the symptoms from the following:

1. Select the month of plantation? April, may, June, July, august or October
2. Select Plant standard position? Normal, Less than normal or Greater than normal
3. Is the plant hailing? Yes or No
4. What is the Temperature of the crop? Normal, Less than normal or Greater than normal
5. Select the crop history? Different lst year, same lst year, same lst two year
6. Select which position of area damaged? Scattered, low areas, upper areas
7. Select severity of the crop? Minor, pot-severe, severe



Soya Bean Expert Sytem

Effected with :	anthracnose
Cure is :	Rotate with a non-legume (non-host) crop. Use pathogen-free seed. Treatment of infected seeds with a fungicide may be beneficial. Fungicidal sprays may be helpful after flowering begins. Soybean varieties differ in their resistance to anthracnose.

Fig.7. Result from the Machine Learning Algorithm

Description:

In this screenshot, the user can see the following.

May be affected with **“**Downy mildew”

Cure Is: Downy mildew can be reduced by planting resistant varieties with resistance to this disease. Delayed planting and foliar fungicides may be beneficial. Spray Dithane M-45 @ 2-2.5 gm/liter.

VII. Conclusions

Bagging Algorithm results improved test set accuracy over single tree classifiers. And also the present system results better compared to the system developed based on the rule based algorithm alone. The implementation of the proposed system also gradually reduces the processing time of rules and gives the solution to the problem with high accuracy. By the thorough interaction with the users and beneficiaries the functionality of the system can be extended further to many more areas.

VIII. References

[1] ERIC BAUER, RON KOHAVI, An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, Kluwer Academic Publishers1998, 3 oct 1997, Boston, Netherlands.

[2] Breiman, Leo. Bagging predictors. Technical Report 421, Department of Statistics, University of California at Berkeley, 1994.

[3] Breiman, L.: Arcing the edge. *Technical report 486, at UC Berkeley, 1996.*

[4] Quinlan, J.R. Induction of Decision Trees. *Machine Learning*, 1: 81-106, 1996.

[5] Quinlan, J. R.: Bagging, boosting and C4.5. *In Proc. of the Fourteenth National Conference on Artificial Intelligence, 1996*

[6] A Comparative Study of Classification Algorithms for Spam Email Data Analysis, Aman Kumar Sharma, IJCST-2010.

[7] ERIC BAUER, RON KOHAVI, an Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, Kluwer Academic Publishers1998, 3 oct 1997, Boston, Netherlands.

[8] L. Breiman, “Pasting small votes for classification in large databases and on-line,” *Machine Learning*, vol. 36, pp. 85–103, 1999.

[9] Breiman. Leo, "Bagging Predictors, in Machine Learning," 1996. P.123-140.

[10] Guo-Zheng Li, and Tian-Yu Liu, “Feature Selection for Bagging of Support Vector Machines,” in *Conf PRICAI 2006*, pp.271–277, 2006.[8] Xuchun Li, Lei Wang, Sung.E, “A study of AdaBoost with SVM based.

[11] Caruana.R and Niculescu-Mizil, “An empirical comparison of supervised learning algorithms”, In Proceedings *of the 23rd international conference on Machine learning*. 2006.

[12] Bendi Venkata Ramana and Prof. M.Surendra Prasad Babu “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis” International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011.

[13] Prof.M.S. Prasad Babu , N.V .Ramana Murty and S.V.N.L .Narayana ”A Web Based Tomato Crop Expert Information System Based On Artificial Intelligence Issn:0975-9646 And Machine Learning Algorithms” International Journal of Computer Science and Information Technologies**,** Vol. 1 (1) , 2010, 6-15.

[14] www.indiakisan.net

[15] Prof. M.S. Prasad Babu, N.V. Ramana Murty, S.V.N.L.Narayana, “A Web Based Tomato Crop Expert Information System Based on Artificial Intelligence and Machine learning algorithms”, *International Journal of Computer Science and Information Technologies, Vol. 1 (1), (ISSN: 0975-9646).*, 2010, pp6-15.

[16] Caruana.R and Niculescu-Mizil, “An empirical comparison of supervised learning algorithms”, In Proceedings *of the 23rd international conference on Machine learning*. 2006.

AUTHOR’S PROFILE

Prof . M. S. Prasad Babu
Professor, Deptt. of CS & SE, A. U. College of Engineering, Andhra University, Visakhapatnam.

Swetha Reddy
M.Tech., Dept of CS & SE, Andhra University, Visakhapatnam, A.P, India.

B. Venkata Ramana,
Assoc. Prof, Dept. ofIT, AITHAM, Tekkali, Srikakulam, A.P, India.

N.V.Ramana Murty
M.Tech., Ph.D. Associate Professor, Dept.of MCA, Research Scholar, Rayalaseema University, Kurnool.