iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 6 **Issue:** V **Month of publication:** May 2018

**DOI:** http://doi.org/10.22214/ijraset.2018.5030

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 6 Issue V, May 2018- Available at www.ijraset.com*

# Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils

Raunak Jahan[1]
[1]*M. Tech (C.S.E) Sam Higginbottom University of Agriculture and Science technology*

***Abstract:** India is predominantly an farming country. Indian soil is classified into two diverse types North Indian Soil and South Indian soil. We have separated it into two basically dissimilar sets are north Indian soil and South Indian Soil. North Indian Soil is the Northern India plains are habitually created of deep alluvial soil. The top soil varies in surface from sand to clay, the larger part being light loam, leaky in texture, simply worked and in nature fruitful. South Indian soil is the Southern India peninsular earth plane is prepared up of hills and river valleys. Hilly tracts are naturally not fitting for farming, some moorland are very warm [Prakash,2015]. The objective of this work is to classify soil types by their characteristics and fertility, predict soil types through classification technique like naïve bayes and also compared this classification technique to among classification technique like zeroR and stacking. This research goal is to judge new data mining techniques apply on soil database for classification of soil and in addition contrast the naïve bayes classifier to another classifier [3]. Soil database is hiered from multiple soil websites like Krishi vigyan kendra, Krishiworld.com, farmer.gov.in and krishijagran.com.*

***Keywords:** Data mining, classification, types of classification, Naïve bayes, soil, validation and evaluation technique, soil database, weka tool description.*

## I. INTRODUCTION

Data mining is the method of learning patterns in huge data sets linking methods at the joint of machine learning, statistics and database systems. It is an vital method where sharp methods are practical to take out data patterns. The generally aim of the data mining procedure is to take out info from a data set and alter it into an logical building for extra use. Data mining is the study step of the KDD (collection, extraction, warehousing, analysis and statistics)**.** Data mining software applications contains a diversity of procedures that have been developed by both corporation and examine centers. These methods have been used in occupation for profit and technical purposes [3].

Farming and natural research studies have used diverse methods of data learn containing, natural trees, statistical machine learning and further analysis method [6]. Data Mining becomes famous in the area of farming for soil classification, wasteland managing, crop and insect managing. In calculated the variety of association technique in data mining and applied into the database of soil science to forecast the imperative relations and provided association rules for unlike soil types in farming. Similarly, farming prediction, disease finding and enhancing the pesticides are analyzed with the use of several data mining methods. In examined the uses of many DM methods for knowledge discovery in farming sector and introduced dissimilar exhibits for knowledge discovery in the form of association Rules, clustering, classification and correlation.

In forecast the soil fertility classes using with classification techniques were naïve Bayes, zeroR and stacking. Data Mining is important to discover the farming related knowledge such as soil fertility, yield prediction and soil erosion. Soil forecast helps for soil cure and crop managing.

Classification algorithms contain finding rules that divide the data into disjoint sets. A set of classification rules are created by such a classification process, which can be used to classify future data [15].

## II. AGRICULTURAL DATA MINING

Data mining software applications contains a diversity of methods that have been developed by both trade and enquiry centers. These methods have been used for occupation, profitable and technical purposes. For example, data mining has been used to inspect soil data sets and found useful classification and designs in the data sets. farming and biological research studies have used varied methods of data study including, natural trees, numerical machine learning and added analysis techniques [3]. Data Mining is important to notice the farming related knowledge such as soil fertility, yield prediction and soil erosion. Soil prediction helps to for soil remedy and crop management [15].

## III. CLASSIFICATION

Classification is a common process linked to categorization, the procedure in which thoughts and objects are familiar, differentiated, and unstated. The three classifier were used to categorize the soil data with Naive bayes algorithm, ZeroR and Stacking which is existing in weka tool. The classification of soil was considered severe to learn because depending upon the productivity class of the soil the area information experts determines which crops should be taken on that exacting soil and which fertilizers should be used [7].

## IV. CLASSIFICATION OF SOIL

Categorization of soil based on the silt, clay and loam with soil color, soil temperature, Ph of soil, soil name and it gives the every information about soil with their properties and fertility. Deals with the orderly classification of soils based on distinctive characteristics as well as criterion that state choices in utilize. Soil classification is a active subject, from the structure of the system itself, to the definitions of classes, and lastly in the application in the field. Soil categorization can be approach from the perception of soil as a matter and soil as a resource [3].

## V. NAIVE BAYES CLASSIFIER

A Naive bayes classifier is one of the classifiers in a family of elegant probabilistic categorization methods in machine learning. It is based on the bayes theorem with freedom features. Each class labels are projected through probability of given instance. Naive bayes classifiers assume that the value of a particular quality is independent of the value of any other quality, given the class variable.

## VI. NAIVE BAYES ALGORITHM

Bayes theorem provides a techniques that we can compute the possibility of a guess given our former knowledge.
Bayes' Theorem is declared as:

P(h|d) = (P(d|h) * P(h)) / P(d) Where
P(h|d) is the possibility of guess h given the data d. This is called the posterior probability.
P(d|h) is the possibility of data d given that the guess h was correct.
P(h) is the possibility of guess h being correct (regardless of the data). This is called the prior probability of h.
P(d) is the possibility of the data (regardless of the guess) [Brownlee,2016] .

## VII. SOIL

Soil is a concoction of minerals, organic matter, gases, liquids, and limitless organisms that mutually maintain life on Earth. Soil is a natural body called the pedosphere which has four imperative functions: It is a medium for plant growth. It is a means of water storage, supply and purification. It is a modifier of Earth's atmosphere. It is a locale for organisms all of which in turn to transform the soil and vegetation. Therefore, India has a huge diversity of soil groups, distinctly dissimilar from one another. India is a country of huge dimensions with varied conditions of geologyrelief and cli mate.

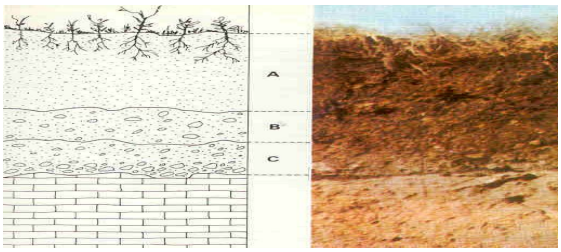


Fig 1:Soil Surface

## VIII. VALIDATION AND EVALUATION TECHNIQUES ARE USED

### *A. Validation Technique*

In this thesis used 10 fold cross validation for estimate result which are averaged. Validation methods in machine learning are used to find the error rate of the ML model, which can be measured as close to the true error rate.

*1)* Partition of dataset into 10 equal parts (fold).

*2)* Grasp out each part in turn.
*3)* Average the outcome.
*4)* Each data point used once for testing , 9 times for training.

*B. Evaluation Technique*

*1) Precision:* In the area of information retrieval, precision is the division of retrieved documents that are relevant to the question.
precision = | { relevant documents } ∩ { retrieved documents } | | { retrieved documents } |

For example, for a book explore on a set of documents, precision is the number of correct outcome separated by the number of all returned results.
Precision takes all retrieved documents into report, but it can also be evaluated at a given cut-off rank, consider only the topmost outcome returned by the system.
Precision is used with recall, the percent of all relevant documents that is returned by the explore. The two process are sometime s used jointly in the F1 Score (or f-measure) to provide a only measurement for a system.

$$\text{Precision} = \frac{tp}{tp + fp}$$

In binary classification, recall is called sensitivity. It can be viewed as the possibility that a relevant document is retrieved by the question. It is slight to achieve recall of 100% by returning all documents in reply to any question. Therefore, recall only is not sufficient but one needs to compute the number of non-relevant documents also, for example by also calculating the precision.

$$\text{Recall} = \frac{tp}{tp + fn}$$

*C. F-measure*

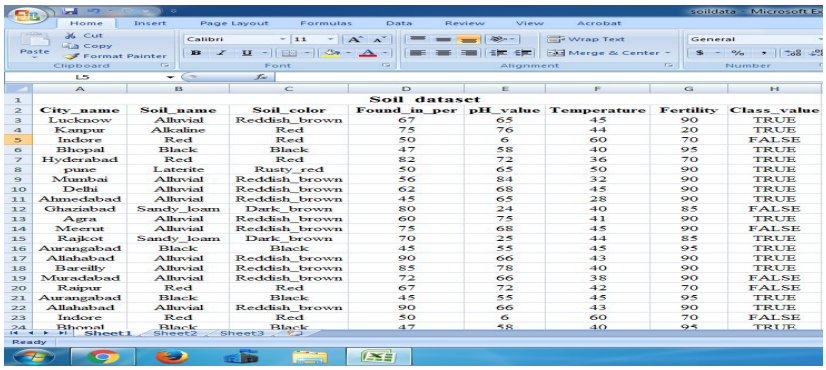
A calculate that combines precision and recall is the harmonic mean of accuracy and recall, the traditional F-measure or balanced F-score.

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

This compute is about the average of the two when they are close, and is more usually the harmonic mean, which, for the case of two numbers, coincides with the square of the geometric mean divided by the arithmetic mean.

## IX. RESULTS AND DISCUSSION

In this thesis, we have collected agricultural soil from multiple agricultural websites like Krishi vigyan kendra and by many soil websites. This soil dataset has been taken by multiple website not only one website. In soil dataset used 50 instances and 8 attributes. We discussed about soil of different states of India through soil's properties and their fertility. We have taken three classification algorithm such as zeroR, stacking and naïve bayes for classification of soil. Naive bayes classifier generates the good performance for this soil data set. As comparing these three algorithms zeroR, stacking resulted in high accuracy.



| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | Soil_dataset | | | | |
| 2 | City_name | Soil_name | Soil_color | Found_in_per | pH_value | Temperature | Fertility | Class_value |
| 3 | Lucknow | Alluvial | Reddish_brown | 67 | 65 | 45 | 90 | TRUE |
| 4 | Kanpur | Alkaline | Red | 75 | 76 | 44 | 20 | TRUE |
| 5 | Indore | Red | Red | 50 | 6 | 60 | 70 | FALSE |
| 6 | Bhopal | Black | Black | 47 | 58 | 40 | 95 | TRUE |
| 7 | Hyderabad | Red | Red | 82 | 72 | 36 | 70 | TRUE |
| 8 | pune | Laterite | Rusty_red | 50 | 65 | 50 | 90 | TRUE |
| 9 | Mumbai | Alluvial | Reddish_brown | 56 | 84 | 32 | 90 | TRUE |
| 10 | Delhi | Alluvial | Reddish_brown | 62 | 68 | 45 | 90 | TRUE |
| 11 | Ahmedabad | Alluvial | Reddish_brown | 45 | 65 | 28 | 90 | TRUE |
| 12 | Ghaziabad | Sandy_loam | Dark_brown | 80 | 24 | 40 | 85 | FALSE |
| 13 | Agra | Alluvial | Reddish_brown | 60 | 75 | 41 | 90 | TRUE |
| 14 | Meerut | Alluvial | Reddish_brown | 75 | 68 | 45 | 90 | FALSE |
| 15 | Rajkot | Sandy_loam | Dark_brown | 70 | 25 | 44 | 85 | TRUE |
| 16 | Aurangabad | Black | Black | 45 | 55 | 45 | 95 | TRUE |
| 17 | Allahabad | Alluvial | Reddish_brown | 90 | 66 | 43 | 90 | TRUE |
| 18 | Bareilly | Alluvial | Reddish_brown | 85 | 78 | 40 | 90 | TRUE |
| 19 | Muradabad | Alluvial | Reddish_brown | 72 | 66 | 38 | 90 | FALSE |
| 20 | Raipur | Red | Red | 67 | 72 | 42 | 70 | FALSE |
| 21 | Aurangabad | Black | Black | 45 | 55 | 45 | 95 | TRUE |
| 22 | Allahabad | Alluvial | Reddish_brown | 90 | 66 | 43 | 90 | TRUE |
| 23 | Indore | Red | Red | 50 | 6 | 60 | 70 | FALSE |
| 24 | Bhopal | Black | Black | 47 | 58 | 40 | 95 | TRUE |

Fig.2: Soil dataset in excel worksheet

These are three threshold curve of naïve bayes, stacking and zeroR classifier in which we can see that clearly naïve bayes gives better result among these classifier because naïve bayes represents more true positive rate than false positive rate.

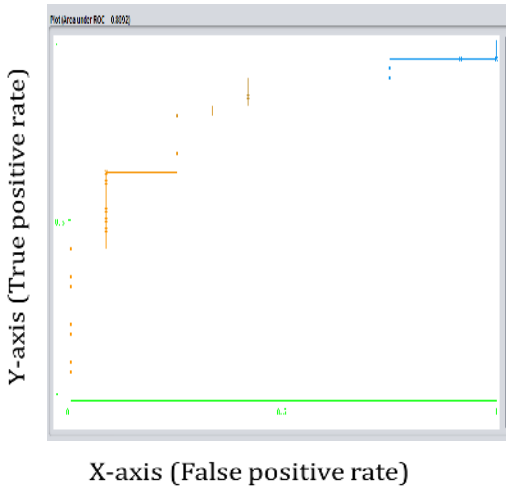



Fig.3: Threshold curve (true class value) of naïve bayes classifier

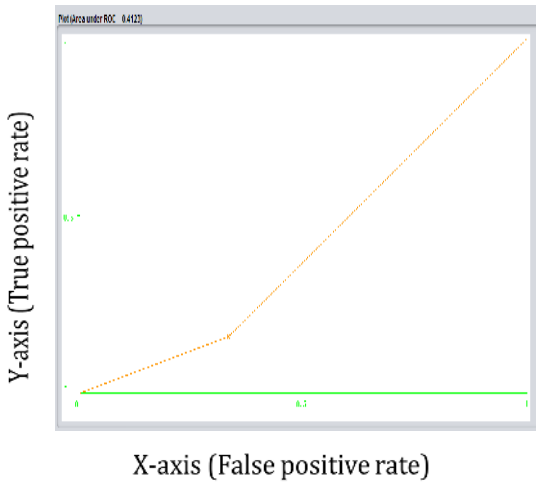
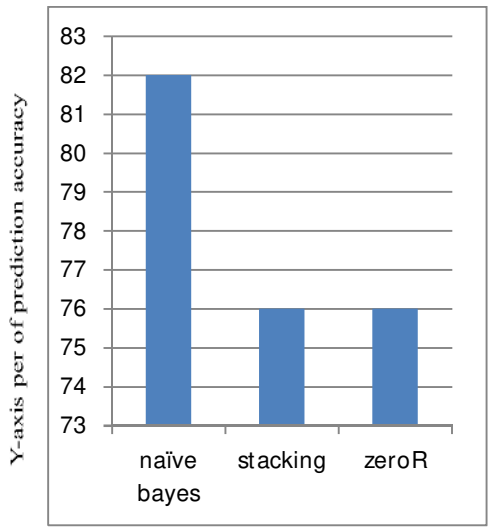



Fig.4:Threshold curve (true class values of zeroR and stacking classifier




Fig;4: Prediction accuracy

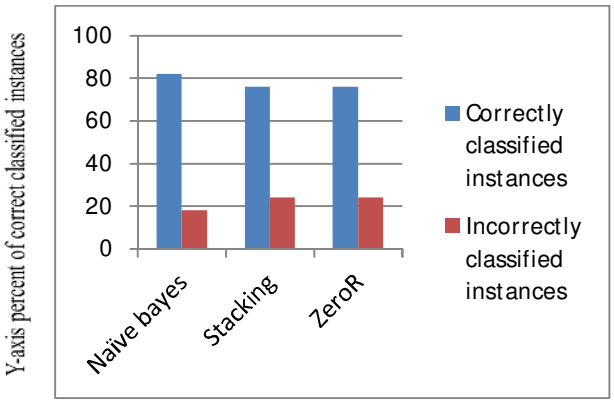



Fig.5: Correct and incorrect classified instances

## X. CONCLUSION

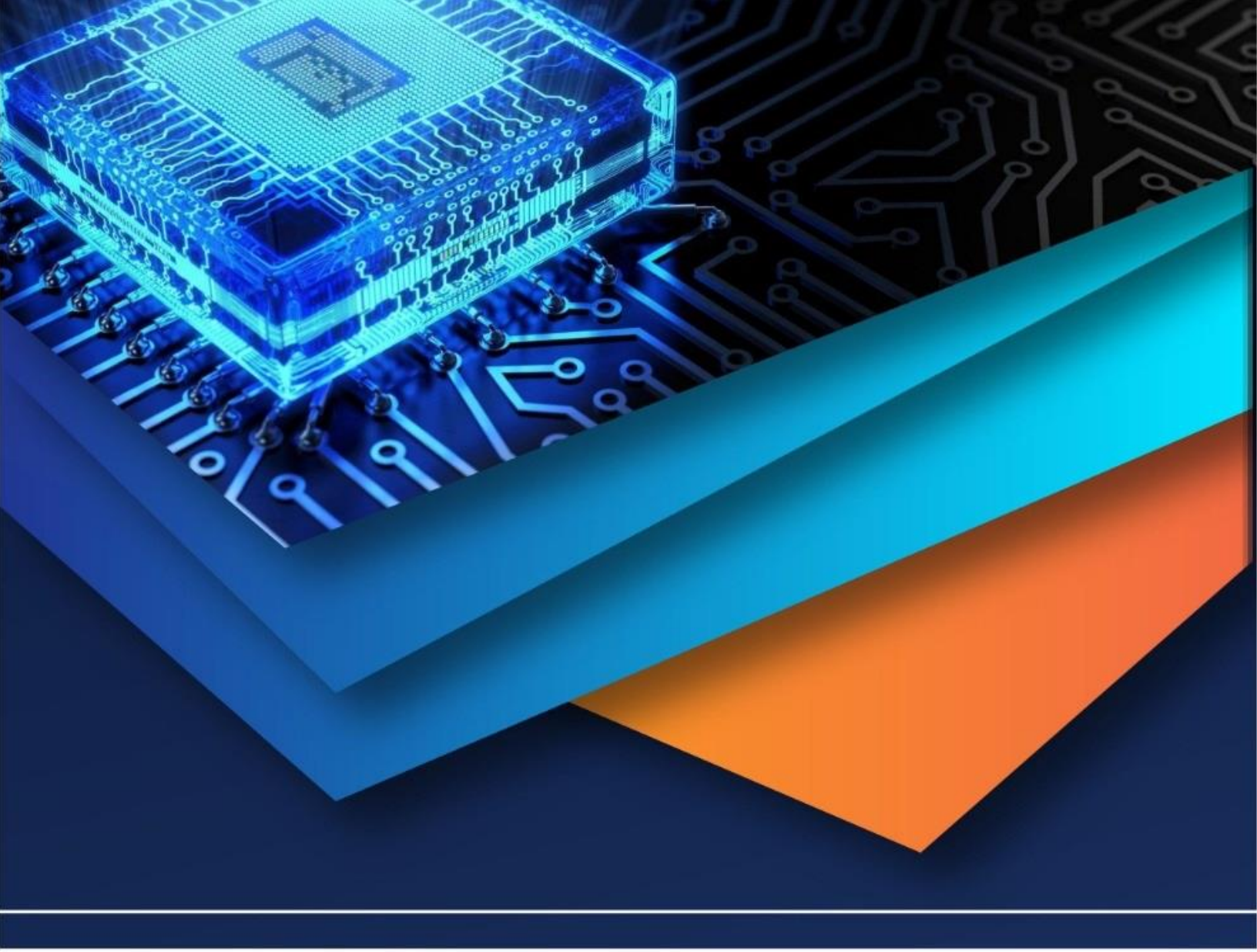
In this paper, the relative study of three algorithms like Naive Bayes, zeroR and stacking is projected. Naive bayes classification algorithm gives superior effect of this dataset and is correctly classified into maximum number of instances comparing with the other two classifier. Naive bayes can be suggested to predict soil types [15]. So we can used this technique in future for classification. This method has found to be quick and simple for classification and generate something new dreams and progress in the field of agriculture.

## XI. FUTURE WORK

In future we have a huge work of agriculture because it is necessary for human beings and now a days every food supplier only wants their profit and makes fatal in the face of food product which goes harmful. We can apply soils in many another tools and check their extra functionalities, properties and characteristic s of soil. So we can give the better solution for better health.Which is important for the human beings.

## REFERENCES

[1] NiteshV.Chawla,KevinW.Bowyer, Lawrence O.Hall & W. Philip Kegelmeyer(2002).SMOTE:Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, Vol 16, 321-357.
[2] D. R.Campbell, S. J. Culley, C.A. McMahon, F.Sellini (2007). An approach for the capture of context-dependent document relationships extracted from Bayesian analysis of users interactions with information, Springer, Vol 10,115-141.
[3] P.Bhargavi & Dr. S. Jyothi (August,2009). Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils, IJCSNS International Journal of Computer Science and Network Security, Vol.9, No.8.
[4] Ron kohavi(2011). Scaling up the accuracy of naïve bayes classifiers: A decision tree hybrid, Research Gate.
[5] Jia WU, Zhihua CAI(2011). Attribute Weighting via Differential Evolution Algorithm for Attribute Weighted Naive Bayes (WNB), Journal of Computational Information Systems, 1672-1679.
[6] Prem Kumar Chandrakar, Sanjay Kumar and Dewashish Mukherjee (2011). Applying Classification Techniques In Data Mining In Agricultural Land Soil,IJCE International Journal Of Computer Engineering, July-December 2011, Vol 3, No 2, pp. 89–95.
[7] Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade & Vahida Attar(2012). Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction [pdf], <https://arxiv.org/pdf/1206.1557>cs.[accessed 2012].
[8] Raorane A.A, Kulkarni R.V (2012). Data Mining: An effective tool for yield estimation in the agricultural sector, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol 1, Issue 2, July-August 2012 ISSN 2278-6856.
[9] S.S.Baskar, L.Arockiam & S.Charles (2013). Applying Data Mining Techniques on Soil Fertility Prediction, International Journal of Computer Applications Technology and Research ,Vol 2, Issue 6, 660 - 662, 2013, ISSN: 2319–8656.
[10] Vivek Narayanan, Ishan Arora, and Arjun Bhatia(2013). Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model, Springer, pp 194-201.
[11] Ajay Kumar Shrivastava & R. N. Panda(2014). Implementation of Apriori Algorithm using WEKA, KIET International Journal of Intelligent Computing and Informatics, Vol. 1, Issue 1.
[12] Vrushali Bhuyar (2014). Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) ,Vol 3, Issue 2, March-April 2014 ISSN 2278-6856.
[13] Hetal Patel, Dharmendra Patel (2014). A Brief survey of Data Mining Techniques Applied to Agricultural Data, International Journal of Computer Applications (0975–8887) ,Vol 95, No. 9, June 2014.
[14] Durga Karthik, K.Vijayarekha (2015). Simple and Quick Classification of Soil for Sunflower Cultivation using Data Mining Algorithm. International Journal of ChemTech Research,Vol.7, No.6, pp 2601-2605.
[15] V.Rajeswari & K.Arunesh(May-2016). Analysing Soil Data using Data Mining Classification Techniques, Indian Journal of Science and Technology, Vol 9(19).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY