VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

Prediction of Crop Yield using Data Mining Approach

In partial fulfilment of the Fourth Year (Semester–VII), Bachelor of Engineering

(B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2017-2018

Project Mentor
Mrs.Priya R.L

Submitted by

Rushika Ramesh Ghadge , D17B/17
Juilee Nitin Kulkarni, D17B/35
Pooja Deepak More, D17B/45
Sachee Milind Nene, D17B/49

(2017-18)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

CERTIFICATE of

Approval

This is to certify that _____of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on "***Prediction of crop yield using data mining approach***" as a part of the coursework of PROJECT-I for Semester-VII under the guidance of ***Mrs.Priya R.L.*** in the year 2017-2018.

_____
Date

_____ _____
Internal Examiner External Examiner

_____ _____
Project Mentor Head of the Department Principal

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Priya R.L** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our
Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to: -

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.

CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

ABSTRA
CT

Agriculture sector of India suffers due to lack of technology solutions as farmers rely on traditional techniques which are less efficient. The use of intelligent systems is more important to increase the productivity. Maximum number of people rely on agriculture for employment. Out of 4.55 crore jobs in Maharashtra, 2.60 crore jobs are related to agriculture and related sectors. As per the 2011 census, out of the total workforce of 45.5 million people in Maharashtra, 26 million were employed in the agriculture and similar sector as cultivators or farm laborers. Total number of registered farmers in the state are 13.6million, and more than 10 million are in the 'small farmers' category, and remaining are in 'marginal farmers' category. So, there is a large majority of small and marginal farmers. Few of the priority areas in this sector includes enhancing agricultural productivity and growth of rural area, improving water resources and irrigation system.

Looking at the current situation of farmers in Maharashtra, we have observed that there is an increase in suicide rate over the years. The reasons behind this includes weather conditions, debt, family issues and frequent change in Indian government norms. Sometimes farmers are not aware about the crop which suits their soil quality, soil nutrients and soil composition. The work proposes to help farmers check the soil quality depending on the analysis done based on data mining approach. Thus, the system focuses on checking the soil quality to predict the crop suitable for cultivation according to their soil type and maximize the crop yield with recommending appropriate fertilizers.

IND
EX

-Chapter 1 : Introduction

1.1. Motivation for the project

As per the statistics of 2016 around 272.82 million farmers dwell in Maharashtra. As per the land use statistics 2013-14, the total geographical area of the country is 328.7 million hectares, of which 141.4 million hectares is the reported net sown area and 200.9 million hectares is the gross cropped area with a cropping intensity of 142 %. The net sown area works out to be
43% of the total geographical area. The net irrigated area is 68.2 million hectares. Despite of this, there has been continuous decline in the share of agriculture to Gross Value Added from
18.2 percent in 2012-13 to 17.0 percent in 2015-16 .With this myriad number of farmers**,** increasing suicide rates and decline in its contribution to GVA we want to help farmers to understand the importance of prior crop prediction, to flourish their basic knowledge about soil quality, understanding location-wise weather constraints, in order to achieve high crop yield through our technology solution.(Annual Report by Department of Agriculture 2016-17)

Table 1.1: Growth in Gross Value Added (at 2012-13 basis prices)

PERIOD	TOTAL GVA	AGRICULTURAL ALLIED GVA (in %)
2012-13	5.4	1.5
2013-14	6.3	4.2
2014-15	7.1	-0.2
2015-16	7.2	1.2

1.2. Problem Definition

Most of the existing system are hardware based which makes them expensive and difficult to maintain. Also, they lack to give accurate results. Some systems suggest crop sequence depending on yield rate and market price. The system proposed tries to overcome these drawbacks and predicts crops by analyzing structured data.

1.3. Relevance of the Project

The project being “Prediction of soil quality using data mining approach” certainly focuses on agricultural aspects. Being a totally software solution, it does not allow

maintenance factor to be considered much. Also, the accuracy level would be high as compared to hardware based solutions as components like soil composition, soil type, pH value, weather conditions all come into picture during the prediction process.

1.4. Methodology used

1.4.1. Data set Collection:

The data sets have been collected and refined based on commonality uses such as soil type, temperature, humidity, soil composition and soil pH. These data sets are entered into the database using MySQL queries. From these parameters name of the crop and net yield rate of the crop can be predicted.

1.4.2. Soil Quality Prediction:

Based on various analyses the parameters pH value and location are taken as input and prediction have been undertaken. For soil quality prediction two algorithms will execute simultaneously. The output given by both algorithms will be compared for accuracy and the most accurate result will be given. The attribute location is used for deriving the type of soil in a particular region such as Alluvial, Loam, Black soil, Clay and Red and the attribute pH value specifies the nature i.e. acidic, basic or neutrality of the soil.

1.4.3. Recommendation of Fertilizers:

Proposed System also recommends fertilizers depending on current soil nutrients and crop predicted. It will compare nutrients needed for predicted crop and current soil nutrients.

1.4.4. Algorithm Used: The proposed system will be implemented using following two algorithms:

1.4.4.1. Supervised learning method: Back Propagation Algorithm

A very common method of teaching artificial neural networks how to perform a given task. BPN able to compute and analyze a large number of data dividing into several small networks and each network can be trained separately along with integration. This algorithm is applied to multilayer feed-forward networks consisting of processing elements with continuous differentiable activation functions.

BPN calculates the accurate result by giving error as feedback till desired output is achieved.

The training of the BPN is done in three stages- the feedforward of the input training pattern, the calculation and back-propagation of the error and updation of weights. The commonly used activation function for BPN is binary sigmoidal and bipolar sigmoidal activation function because of characteristics like continuity, differentiability and non-decreasing monotony.

1.4.4.2. Unsupervised learning method: Kohonen Self Organizing Maps
Self-organizing neural networks are used to cluster input patterns into groups of similar patterns. They are called maps because they effectively map weights to input data and assume a topological structure. The Kohonen network is probably the best example, because it's simple and has concepts of self-organization and unsupervised learning easily. It gives the best result with no human intervention and needs to know little about the characteristics of the input data. Each input has a weight and the different input patterns are shown to all neurons simultaneously.

The structure of a self-organizing map involves *m* cluster units, arranged in either a one- or two-dimensional array, with vectors of *n* input signals. Each cluster is defined by weights. Different input patterns are compared with the cluster and the best match is associated with the cluster. The comparison is usually done on the basis of minimum Euclidean distance
The formation of a map occurs in two stages:

1. The initial formation of the correct order
2. and the final convergence

The second stage takes much longer, and usually occurs when the learning rate gets smaller.

1.4.4.3. Fuzzy Logic
Fuzzy logic is to convey the uncertainties in the membership function. The functions are usually constructed through numerical data or range of classes. Uncertainty lies in the shape of the membership and thus cannot represent a crisp value. Fuzzy logic is used to develop a fuzzy control system which is useful in obtaining crisp output from a fuzzy input. The crisp output values are calculated using fuzzy rules applied in an inference engine using defuzzification methods.

Fuzzy logic system has a fuzzifier, an inference engine and a defuzzifier. Fuzzy logic can be used as an interpretation model for the properties of neural networks, as well as for giving a more precise description of their performance. Fuzzy logic can also be used to specify networks directly without having to apply a learning algorithm. It is straightforward to formulate a set of fuzzy rules for this task, but it is not immediately obvious how to build a network to do the same nor how to train it. Fuzzy logic is now being used in many products of industrial and consumer electronics for which a good control system is sufficient and where the question of optimal control does not necessarily arise.

Chapter 2: Literature Survey

2.1. Papers

Pape

r 1:

Agriculture Analysis for Next Generation High Tech Farming in Data

Min
ing

a. Abstract of the research paper

Agricultural researchers and farmers deploy sensors at their remote agricultural-fields to obtain the data of temperature, humidity, soil moisture and so on. Automatic collection of those data greatly helps their analytical works. Currently, they are relying on network providers, for example, cellular phone network to achieve such automatic collection from their remote sites, which is not feasible for most of the farmers due to operational cost. Our review and meta-analysis of yield data comparing organic and conventional agriculture showed that currently organic yields of individual crops are on average 80% of conventional yields. The analysis of 362 datasets also showed a high variation of the yield gap of organic agriculture (standard deviation 21%). Some of this variation seems systematic. Relative yields differed between crops with e.g. soybean, some other pulses, rice and corn scoring higher than 80% and wheat, barley and potato scoring lower than 80%. Most regions have relative yields fairly close to the overall average.

b. Inference drawn from the paper

Agriculture is the most significant application area particularly in the developing countries like India. Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way. Project technique plays a crucial role for decision making on several issues related to agriculture field. The systems must be compared at rotation level using a unit (e.g. grain equivalents) that allows for comparing farming systems with different crop compositions. It discusses about the role of data mining in the agriculture field and their related work by several authors in context to agriculture domain. It also discusses on different data mining applications in solving the different agricultural problems. This paper integrates the work of various authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field.

Pape
r 2:
A Data-Driven Approach to Soil Moisture Collection and Prediction a. Abstract of the research paper

Agriculture has been one of the most under-investigated areas in technology, and the
development of Precision Agriculture (PA) is still in its early stages. This paper proposes a data-driven methodology on building PA solutions for collection and data modelling systems. Soil moisture, a key factor in the crop growth cycle, is selected as an example to demonstrate the effectiveness of our data-driven approach. On the collection side, a reactive wireless sensor node is developed that aims to capture the dynamics of soil moisture using MicaZ mote and VH400 soil moisture sensor. The prototyped device is tested on field soil to demonstrate its functionality and the responsiveness of the sensors. On the data analysis side, a unique, site-specific soil moisture prediction framework is built on top of models generated by the machine learning techniques SVM (support vector machine) and RVM (relevance vector machine). The framework predicts soil moisture n days ahead based on the same soil and environmental attributes that can be collected by our sensor node. Due to the large data

size required by the machine learning algorithms, our framework is evaluated under the Illinois historical data, not field collected sensor data. It achieves low error rates (15%) and high correlations (95%) between predicted values and actual values across 9 different sites when forecasting soil moisture about 2 weeks ahead. Also, it is shown that the prediction outputs can remain accurate over a long period of time (one year) when reliable data are fed to the model every 45 days.

b. Inference drawn from the paper

Most of the work in this paper focuses on prototyping a reactive sensor node; the performance of the wireless sensor network when deployed at large scale in real world is not tested. Battery life, unit cost and long-term reliability play important roles in sensor deployment and should be addressed in future. On the prediction side, in our prediction experiment, we assume the forecasting data are error-free. Applying real, noise-included weather forecasting data onto models can be done in future work. Lastly, more advanced techniques in machine learning can be explored to reduce the amount of data required to achieve a good model. Currently, obtaining a fairly good model requires at least one year of data. Methods that include spatial factors into the modelling and achieve a fairly good model with less data are potential areas for future research.

Pap er 3: Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks a. Abstract of the research paper

The objective of this paper is to analysis of main soil properties such as organic matter, essential plant nutrients, micronutrient that affects the growth of crops and find out the suitable relationship percentage among those properties using Supervised Learning, Back Propagation Neural Network. Although these parameters can be measured directly, their measurement is difficult and expensive. Back Propagation Networks(BPN) are trained with reference crops' growth properties available nutrient status and its ability to provide nutrients out of its own reserves and through external applications for crop production in both cases, BPN will find and suggest the correct correlation percentage among those properties. This machine learning system is divided into three steps, first sampling (Different soil with same number of properties with different parameters) second Back Propagation Algorithm and third Weight updating. The performance of the Back Propagation Neural network model will be evaluated using a test data set. Results will show that artificial neural network with certain number of neurons in hidden layer had better performance in predicting soil properties than multivariate regression. In conclusion, the result of this study showed that training is very important in increasing the model accuracy of one region and result in the form of a guide to recognizing soil properties relevant to plant growth and protection.

b. Inference drawn from the paper

Instead of random values if real value of soil properties of examining field are taken and if they are compared with standard (desired or target) data set, then above algorithm and code will give the accurate updation needed for soil Compost. Results will show that the proposed correction method enables improving significantly Soil properties prediction accuracy and performs better than traditionally used methods consisting in automatic weight updation processing.

Pap er 4: Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach

a. Abstract of the research paper

Agriculture sector is backbone of Indian Economy. However, Agriculture sector in India is facing severe problem of maximising the crop productivity. Farmers lack in basic knowledge of nutrient content of soil, selection of crop best suited for soil and they also lack in efficient method of prediction of crop well in advance so that appropriate methods can be used to improve crop productivity and to make arrangements for storage, marketing well before harvest. This work presents an approach which uses different Machine Learning techniques
in order to predict
the category of the yield based on macronutrients and micronutrients status in dataset. The dataset considered for the crop yield prediction was obtained from Krishi Bhawan (Talab-Tillo) Jammu. The parameters present in the data are Macro-Nutrients (ph, Oc, Ec, N, P, K, S) and Micronutrients (Zn, Fe, Mn, Cu) present in samples collected from different regions of Jammu District. After analysis Machine learning algorithms are applied to predict the
category of yield. The category, thus predicted will specify the yield of crops. The problem of predicting the crop yield is formulated as Classification where different classifier algorithms are used.

b. Inference drawn from the paper

In this work the Classification of Soils and prediction of yield has been done on the basis of Soil Composition parameters which is one of subsets for Rice Yield Prediction. This Future work can be proposed by considering other important subsets which are Rain, Humidity, Temperature Precipitation. Further if individual farmer's production data is available it can be used as Response parameter to Predict actual production based on input parameters. Additional Machine Learning Techniques like SVM, Rule Based Induction, can be used.

Paper 5:
Using Deep Learning for Image-Based Plant Disease

Detection a. Abstract of the research paper

Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, train a deep convolutional neural network to identify
14 crop species and 26 diseases. The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

b. Inference drawn from the paper

Uses deep convolutional neural network architecture, which trains a model on images of plant leaves for classifying both crop species and the presence and identity of disease on images with 99.35% accuracy.

2.2. Patent search

Patent 1:
a. Title of the patent and year of the patent:

Crop yield prediction-2005

b. Summary of the patent:

The invention relates to a crop transpiration prediction method based on an improved extreme learning machine. First, soil environment data and meteorological data of a farmland are collected and normalized to obtain a training set; then, the training set is adopted to train an extreme learning machine network and improve an extreme learning machine; last,
normalized data collected again is input into the improved extreme learning machine, and the improved extreme learning machine outputs crop transpiration obtained through prediction. Extreme learning machine improvement mainly comprises the steps that 1, a function based on waveform superposition is adopted to serve as an activation function for a hidden layer of the extreme learning machine; 2, a particle swarm optimization algorithm is adopted to optimize an input weight value and a threshold value between a network input layer and the hidden layer of the extreme learning machine. Through the prediction method, the prediction precision of crop transpiration is improved, prediction time loss is reduced, and meanwhile generalization performance and prediction stability of the traditional extreme learning machine network are improved.

c. Link 1. European Patent:

https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=0&ND=3&adj
acent=true&locale=en_EP&FT=D&date=20170510&CC=CN&NR=106651012A&KC=A#

Link 2. US patent: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=crop.TI.&s2=prediction.ABTX.&OS
=TTL/crop+AND+ABST/prediction&RS=TTL/crop+AND+ABST/prediction

Patent 2:
a. Title of the patent and year of the patent:

Crop transpiration prediction method based on improved extreme learning machine-2017

b. Summary of the patent:

The invention relates to a crop transpiration prediction method based on an improved extreme learning machine. First, soil environment data and meteorological data of a farmland are collected and normalized to obtain a training set; then, the training set is adopted to train an extreme learning machine network and improve an extreme learning machine; last,
normalized data collected again is input into the improved extreme learning machine, and the improved extreme learning machine outputs crop transpiration obtained through prediction. Extreme learning machine improvement mainly comprises the steps that 1, a

function based on waveform superposition is adopted to serve as an activation function for a hidden layer of the extreme learning machine; 2, a particle swarm optimization algorithm is adopted to optimize an input weight value and a threshold value between a network input layer and the hidden layer of the extreme learning machine. Through the prediction method, the prediction precision of crop transpiration is improved, prediction time loss is reduced, and meanwhile generalization performance and prediction stability of the traditional extreme learning machine network are improved.

c. Link 1. European Patent:

https://worldwide.espacenet.com/publicationDetails/biblio?II=1&ND=3&adjacent=true&loca le=en_EP&FT=D&date=20170510&CC=CN&NR=106651012A&KC=A#

Link 2. US patent: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=crop.TI.&s2=prediction.ABTX.&OS =TTL/crop+AND+ABST/prediction&RS=TTL/crop+AND+ABST/prediction

Patent 3:

a. Title of the patent and year of the patent:

Crop yield prediction using piecewise linear regression with a break-point and weather and agricultural parameters -2010

b. Summary of the patent

Crop yield may be assessed and predicted using a piecewise linear regression method with break point and various weather and agricultural parameters, such as NDVI, surface parameters (soil moisture and surface temperature) and rainfall data. These parameters may help aid in estimating and predicting crop conditions. The overall crop production environment can include inherent sources of heterogeneity and their nonlinear behaviour. A non-linear multivariate optimization method may be used to derive an empirical crop yield *prediction* equation. Quasi-Newton method may be used in optimization for minimizing inconsistencies and errors in yield *prediction*. Minimization of least square loss function through iterative convergence of pre-defined empirical equation can be based on piecewise linear regression method with break point. This nonlinear method can achieve acceptable lower residual values with predicted values very close to the observed values. The present invention can be modified and tailored for different crops worldwide.

c. Link 1. US patent:

http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=5&f=G&l=50&co1=AND&d=PTXT&s1=%22Crop+yield+prediction%22&OS =%22Crop+yield+prediction%22&RS=%22Crop+yield+prediction%22

Chapter 3: Requirements for the proposed system

The requirements for the project titled "Prediction of crop yield using data mining approach " comprises of: -

3.1 Functional Requirements

1. **Registration and Login**: Farmers can register in the system where he/she can add location and other important details and login accordingly.
2. **Location Tracking**: The android application can track location of farmer and depending on location we can categorise the type of soil and add the data in our database.
3. **Weather forecasts** (seasonal rains, droughts, famine, checking quality of soil): Detecting the weather conditions on a weekly basis and suggesting the farmers about whether the cultivation of the crops should be done within that week.
4. Depending on the rainfall and soil quality accordingly **crops** that can be grown in that particular soil will be **predicted**.
5. Nearby **water resource availability** will be notified to the farmers using GPS functionality.
6. **Display Government schemes**: Whenever government updates any scheme related to discounts on crop seeds etc. application will notify the user about it.

3.2. Non-Functional Requirements

1. **Availability**: Farmers can access the system at any time and view weather forecasts, etc.
2. **Security**: The proposed system's data will be encrypted. Every farmer will have an account.
3. **User Friendly**: Multilingual support for Indian languages such as: Hindi, Marathi, Tamil, Telugu, Kannada, Punjabi and simple User Interface.
4. **Reliability**: The proposed system supports multiple users at a time and servers wont crash even if the load increases.

3.3. Constraints

Scope lies only in Maharashtra.

3.4. Hardware & Software Requirements

3.4.1 Hardware requirements

Mobile Device (Android OS)

Computer:

Processor - dual core @ 2.4 GHz (i5 or i7 Intel processor or equivalent AMD)

RAM - 4 GB / 8 GB

Hard Drive - 320 GB 5400 RPM hard drive

OS: Windows, Linux, MAC

3.4.2 Software requirements

Android Studio
R, PHP
XAMPP, NGNIX
Google prediction API (google play services)

3.5 System Block Diagram

Chapter 4: Proposed Design

4.1 System Design / Conceptual Design (Architectural)

4.2 Detailed Design

4.2.a Data Flow Diagram (Level 0,1,2)

Level 0

Leve
l 1

Leve
l 2

4.2.b Flowchart for the proposed system

4.2.c ER Diagram

4.3. Project Scheduling & Tracking using Timeline / Gantt Chart

Chapter 5. Proposed Results and Discussions

An Android application for users which predicts the crops based on users inputs and recommends fertilizers and detects crop diseases.

The system must support storing and retrieving of user(farmer) details, crop details and soil composition details which will later be accessed by users who log into the system through a user interface that allows privileged users such as administrators to carry out tasks such as adding/editing crop details, adding/editing soil composition details and creating/editing farmer details. It supports viewing functionality in order to allow normal and privileged users to view details of a given entity from the system database.

There will also be a provision to know about different beneficiary schemes like Pradhan Mantri Krishi Sinchayee Yojana(PMKSY), Agriculture credit, Interest Subvention Scheme, Kisan Credit Card, Crop Insurance, etc for the farmers provided by the Government. It will be displayed in the form of a notification which will give updates regularly.

The farmers can take an advantage of the multilingual facility. The information would be displayed in regional languages like Hindi, Marathi, Tamil, Telugu, Kannada, Punjabi etc.

The system will be available anytime anywhere.

The system must have a functionality that produces summary reports from analyses and calculations. The system should be able to run without any errors on popular technical environments (e.g.: UNIX, Windows, Mac OS/2, etc).

Chapter 6. Plan of action for the next semester

a. Work done till date (Phase I):

Finalised project synopsis which includes problem statement, scope of project, functional and non-functional requirements, software specifications and hardware specifications.

Pros of current system over already existing system.

Designed system architecture and UML diagrams for proposed system.

Finalized the algorithms to be used in the system.

b. Plan of action for Phase II:

Implementation of project including GUI, backend, coding, code fixation.

Preparing test scripts, executing text cases and deployment.

Check the system in live environment and obtain feedback from users.

Chapter 7. Conclusions

The system uses supervised and unsupervised Machine learning algorithms and gives best result based on accuracy. The results of the two algorithms:Kohonen SOM and Backpropagation Network will be compared and the one giving the best and accurate output will be delivered to the end users.Kohonen SOM gives best result with no human intervention and little needs to be known about the characteristics of the input data.BPN calculates the accurate result by giving error as feedback till desired output is achieved.Thus the system will improve crop prediction significantly by giving accurate results along with fertilisers recommendations.
The system can be enhanced further to add following functionality :

1. Crop diseases detection using Image Processing where users can upload picture of diseased crop and get pesticides recommendations.

2. Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

Chapter 8. References

[1] P.Vinciya, Dr. A. Valarmathi, “Agriculture Analysis for Next Generation High Tech Farming in Data Mining”*IJARCSSE*, vol. 6, Issue 5, 2016.

[2] Shivnath Ghosh,Santanu Koley, “Machine Learning for Soil Fertility and Plant N utrient Management using Back Propagation Neural Networks”*IJRITCC*, vol. 2, Issue 2,292-297,2014.

[3] Zhihao Hong,Z. Kalbarczyk,R. K. Iyer, “A Data-Driven Approach to Soil Moisture Collection and Prediction”*IEEE Xplore*, vol. 2, Issue 2,292-297,2016.

[4]Sabri Arik, Tingwen Huang, Weng Kin Lai, Qingshan Liu , “Soil Property Prediction: An
Extreme Learning Machine Approach”*Springer*, vol. 3, Issue 4,666-680,2015.

[5]Vaneesbeer Singh,Abid Sarwar, “Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach”*IJARCSE*, vol. 5, Issue 8,2017.

[6]Okalebo et al. (2002). Laboratory Methods of Soil and Plant Analysis: A working Manual. Second edition TSBF-CIAT and SACRED Africa: Nairobi, Kenya.

[7]Wikipedia. (2009). Soil Testing. Last modified March 1st 2009. Retrieved on March 4th
2009 from http://www.wikipedia.com/soil_testing.

[8]Echochem Online. (2009). Soil Health and Crop yields. Last modified January 28th 2009.
Retrieved on March 4th 2009 from http://ecochem.com/healthy_soil.html .

[9]Food and Agricultural Organization. (2006). The state of Agricultural Commodity Markets. 37-39.

[10]Department of Agriculture ,
http://agricoop.nic.in/sites/default/files/Annual_rpt_201617_E.
pdf

Chapter 9. Appendix

9.1. List of Figures

9.2. List of Tables

9.3 Paper Publications

9.3.1. Draft of the paper published

Prediction of Crop Yield using Data Mining Approach

Rushika Ghadge Juilee Kulkarni Pooja More
Sachee Nene rushika.ghadge@ves.ac.in juilee.kulkarni@ves.ac.in
pooja.more@ves.ac.in sachee.nene@ves.ac.in
Priya R L
priya.rl@ves.ac.in

Computer Engineering
VESIT, Mumbai, India

***Abstract*—Looking at the current situation faced by farmers in Maharashtra, we have observed that there is an increase in suicide rate over the years.The reasons behind this includes weather conditions, debt, family issues and frequent change in indian government norms. Sometimes farmers are not aware about the crop which suits their soil quality, soil nutrients and soil composition. The work proposes to help farmers check the soil quality depending on the analysis done based on data mining approach. Thus the system focuses on checking the soil quality to predict the crop suitable for cultivation according to their soil type and maximize the crop yield with recommending appropriate fertilizers .**

Keywords—Kohonen's SOM(Self-Organizing Map), BPN(Back-Propagation Neural Networks), API(Application Programming Interface)

I. INTRODUCTION

[10]As per the statistics of 2016 around 272.82 million farmers dwell in Maharashtra.With this myriad number of farmers and increasing suicide rates , we want to help farmers to understand the importance of prior crop prediction, to flourish their basic knowledge about soil quality, understanding location-wise weather constraints, in order to achieve high crop yield through our technology solution.

Most of the existing system are hardware based which makes them expensive and difficult to maintain.Also they lack to give accurate results.Some systems suggest crop sequence depending on yield rate and market price.The system proposed tries to overcome these drawbacks and predicts crops by analyzing structured data.

[8]The project being "Prediction of soil quality using data mining approach" certainly focuses on agricultural aspects. Being a totally software solution, it does not allow maintenance factor to be considered much. Also the accuracy level would be high as compared to hardware based solutions, because components like soil composition, soil type, pH value, weather conditions all come into picture during the prediction process.

II. LITERATURE SURVEY

[9]Agriculture sector plays a major role in indian economy, as 70 percent households in india depends purely on this field. Agriculture in India

contributes to about 17% of Gross Value Added as of 2015-16.But there is a continuous decline in agriculture's contribution to Gross Value Added.Food is essential for life and we depend on agricultural outputs, so farmers play a very important role.
The following comparison is shown below :

The study in [1] used Multiple Linear Regression(MLR) technique for crop analysis.Decision tree algorithm and Classification is used to perform analysis of over 362 datasets and provide result.The training dataset here is classified into as organic,inorganic and real estate for predicting the type of soil.Results computed by this system are accurate as well as reliable.

The study in [2] fed data to a Back Propagation Network to evaluate the test data set.Back Propagation Network uses a hidden layer which helps in better performance in predicting soil properties.Back Propagation Network here,is employed to develop a self trained function to predict soil properties with parameters.This gives more accuracy and performs better than the traditionally used methods ,however,sometimes the system becomes slow and inconsistency is seen in the output.

In [3] two regression supervised machine learning methods are used:Support Vector Machine(SVM) and Relevance Vector Machine(RVM) to show effectiveness in soil quality prediction.a smart wireless device for sensing soil moisture and meteorological data.The wireless device gives an error rate of 15% and 95% accuracy. However,it has not been tested for real time data.

The paper [4] involves a check for Soil Fertility and Plant Nutrient by using back propagation algorithm.The results are accurate and enables improvement in soil properties.It performs better as compared to traditional methods.However, system is slow inefficient and not stable.

According to paper [5], three methods are used which includes Decision tree, Naive Bayes Classifier, and KNN Classifier which analyses soil and predicts crop yield, However rule based induction and SVM can be used for more accuracy as results are not accurate.

III. Proposed System

The system aims to help farmers to cultivate proper crop for better yield production.To be precise and accurate in predicting crops, the project analyze the nutrients present in the soil and the crop productivity based on location.It can be achieved using unsupervised and supervised learning algorithms, like Kohonen Self Organizing Map (Kohenon's SOM) and BPN (Back Propagation Network). Dataset will then trained by learning networks. It compares the accuracy obtained by different network learning techniques and the most accurate result will be delivered to the end user. Along with this, the end user is provided with proper recommendations about fertilizers suitable for every particular crop.

IV. System Architecture

Fig.1. Block Diagram

[7]The proposed system will check soil quality and predict the cop yield accordingly along with it provide fertiliser recommendation if needed depending upon the quality of soil.

The functionality of the architecture(Fig.1.) is as follows : The system takes inputs pH value (based on percentage of nutrient) and location from the user. Result processing is done by two controllers.

Location is used as an input to controller 1, along with the use of third party applications like APIs for weather and temperature,type of soil, nutrient value of the soil in that region,amount of rainfall in the region , soil composition can be determined.

[6]pH value is given as an input to controller 2, from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen(N), Phosphorous(P), Potassium(K), Sulphur (S), Magnesium (Mg), Calcium (Ca),

Iron (Fe), Manganese, Boron and Zinc and Organic matter can be obtained.
The result of the controller 1 and controller 2 are compared with a predefined “nutrients” data store. These compared results are supplied to controller 3 wherein the combination of the above results and the predefined data set present in the crop data store is compared.
Finally, the results are displayed in the form of bar graphs along with accuracy percentage.

Fig.2.Modular Diagram

V. CONCLUSION

The system uses supervised and unsupervised Machine learning algorithms and gives best result based on accuracy. The results of the two algorithms will be compared and the one giving the best and accurate output will be selected.Thus the system will help reduce the difficulties faced by the farmers and stop them from attempting suicides.It will act as a medium to provide the farmers efficient information required to get high yield and thus maximize profits which in turn will reduce the suicide rates and lessen his difficulties.

VI. FUTURE SCOPE

The system can be enhanced further to add following functionality :

1. Crop diseases detection using Image Processing where users can upload picture of diseased crop and get pesticides recommendations.
2. Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

VIII. REFERENCES

[1] P.Vinciya, Dr. A. Valarmathi, “Agriculture Analysis for Next Generation High Tech Farming in Data Mining”*IJARCSSE*, vol. 6, Issue 5, 2016.

[2] Shivnath Ghosh,Santanu Koley, "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks"*IJRITCC*, vol. 2, Issue 2,292-297,2014.

[3] Zhihao Hong,Z. Kalbarczyk,R. K. Iyer, "A Data-Driven Approach to Soil Moisture Collection and Prediction"*IEEE Xplore*, vol. 2, Issue 2,292-297,2016.

[4]Sabri Arik, Tingwen Huang, Weng Kin Lai, Qingshan Liu , "Soil Property Prediction: An Extreme Learning Machine Approach"*Springer*, vol. 3, Issue 4,666-680,2015.

[5]Vaneesbeer Singh,Abid Sarwar, "Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach"*IJARCSE*, vol. 5, Issue 8,2017.

[6]Okalebo et al. (2002). Laboratory Methods of Soil and Plant Analysis: A working Manual. Second edition TSBF-CIAT and SACRED Africa: Nairobi, Kenya.

[7]Wikipedia. (2009). Soil Testing. Last modified March 1st 2009.Retrieved on March 4th 2009 from http://www.wikipedia.com/soil_testing.

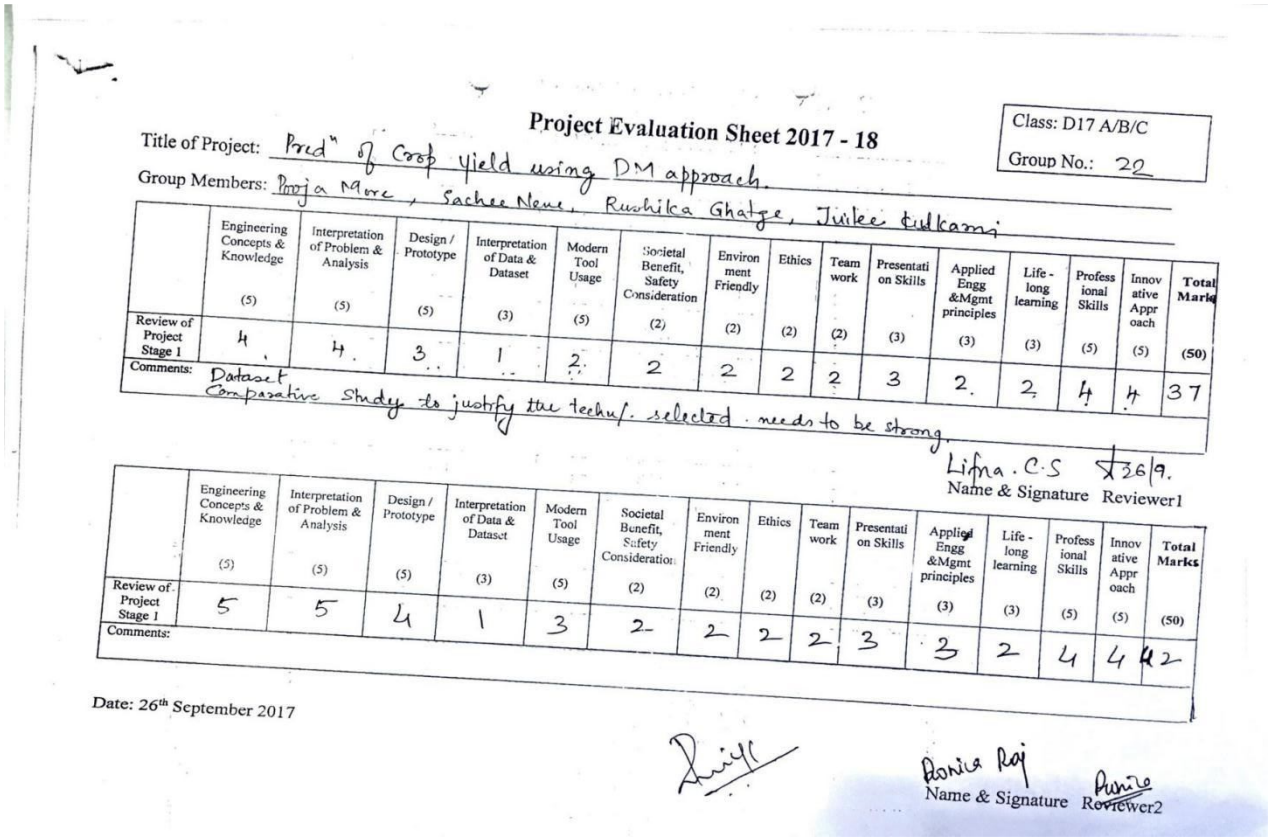
[8]Echochem Online. (2009). Soil Health and Crop yields. Last modified January 28th 2009.Retrieved on March 4th 2009 from http://ecochem.com/healthy_soil.html .

[9]Food and Agricultural Organization. (2006). The state of Agricultural Commodity Markets. 37-39.

[10]http://agricoop.nic.in/sites/default/files/Annual_rpt_201617_E.pdf

9.3.2. Plagiarism report of the paper published /draft

9.3.3. Xerox of project review sheet



Project Evaluation Sheet 2017 - 18

Class: D17 A/B/C

Group No.: 22

Title of Project: Pred" of Crop Yield using DM approach.

Group Members: Pooja More, Sachee Nene, Rushika Ghatge, Jui Kee Kulkarni

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	3	1	2	2	2	2	2	3	2	2	4	4	37

Comments: Dataset Comparative Study to justify the techn. selected needs to be strong.

Lifna. C.S 26/9.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	5	5	4	1	3	2	2	2	2	3	3	2	4	4	42

Comments:

Date: 26th September 2017

Monica Raj

Name & Signature Reviewer2