Abstract—With the impact of climate change in India,

majority of the agricultural crops are being badly affected

interms of their performance over a period of last two

decades. Predicting the crop yield well ahead of its harvest

would help the policy makers and farmers for taking

appropriate measures for marketing and storage. Such

predictions will also help the associated industries for

planning the logistics of their business. Several methods of

predicting and modeling crop yields have been developed in

the past with varying rateof success, as these don’t take

into account characteristicsoftheweather, a n d aremostly

empirical. In the present study a software tool named

‘Crop Advisor’ has been developed as an user friendly web

page for predicting the influence of climatic parameters on

the crop yields.C4.5 algorithm is used to find out the most

influencing climatic parameter on the crop yields of selected

crops in selected districts of Madhya Pradesh. This software

provides an indication of relative influence of different

climatic parameters on the crop yield, other agro-input

parameters responsible for crop yield are not considered in

this tool, since, application of these input parameters varies

with individual fields in space and time.

Key words: Climate, agricultural productivity, C4.5

alogarithm, prediction

I.INTRODUCTION

Crop production is a complex phenomenon that

is influenced by agro-climatic input parameters.

Agriculture input parameters varies from field to

field and farmer to farmer. Collecting such

information on a larger area is a daunting task.

However, the climatic information collected in India

at every 1sq.m area in different parts of the district

are tabulated by Indian Meteorological Department.

The huge such data sets can be used for predicting

their influence on major crops of that particular

district or place. There are different forecasting

methodologies developed and evaluated by the

researchers all over the world in the field of

agriculture or associated sciences. Some of the such

studies are : Agricultural researchers in Pakistan

have shown that attempts of crop yield maximization

through pro-pesticide state policies have led to a

dangerously high pesticide usage. These studies

have reported negative correlation between pesticide

usage and crop yield[1]. In their study they have

shown that how data mining integrated agricultural

data including pest scouting, pesticide usage and

meteorological data are useful for optimization of

pesticide usage. Thematic information related to

agriculture which has spatial attributes was reported

in one of the study[6]. Their study aimed at

discerning trends in agriculture production with

references to the availability of inputs. K- means

method was used to perform forecasts of the

pollution in the atmosphere [4], the k nearest

neighbor was applied for simulating daily

precipitations and other weather variables [11], and

different possible changes of the weather scenarios

are analyzed using SVMs[13]. Data mining

techniques are often used to study soil

characteristics. As an example, the k-means

approach is used for classifying soils in combination

with GPS-based technologies [14]. Apples were

checked using different approaches before sending

them to the market.[9], uses a k-means approach to

analyze color images of fruits as they run on

conveyor belts. [12] uses X-ray images of apples to

monitor the presence of water cores, and a neural

network is trained for discriminating between good

and bad apples. Spatial data mining introduced

especially decision tree algorithm applying to

agriculture land grading[15]. He combined spatial

data mining techniques with expert system

techniques and applied them to establish an

intelligent agriculture land grading information

system. The author adopted decision tree C4.5

algorithm and implement with Mo2.0 and VC++6.0

to build agriculture land grading expert system. The

study showed the advantages of this method in

addressing problems in land grading. A decision

tree classifier for agriculture data was proposed

[5].This new classifier uses new data expression and

can deal with both complete data and in complete

data. In the experiment,10-fold cross validatio

Abstract—With the impact of climate change in India,

majority of the agricultural crops are being badly affected

interms of their performance over a period of last two

decades. Predicting the crop yield well ahead of its harvest

would help the policy makers and farmers for taking

appropriate measures for marketing and storage. Such

predictions will also help the associated industries for

planning the logistics of their business. Several methods of

predicting and modeling crop yields have been developed in

the past with varying rateof success, as these don’t take

into account characteristicsoftheweather, a n d aremostly

empirical. In the present study a software tool named

‘Crop Advisor’ has been developed as an user friendly web

page for predicting the influence of climatic parameters on

the crop yields.C4.5 algorithm is used to find out the most

influencing climatic parameter on the crop yields of selected

crops in selected districts of Madhya Pradesh. This software

provides an indication of relative influence of different

climatic parameters on the crop yield, other agro-input

parameters responsible for crop yield are not considered in

this tool, since, application of these input parameters varies

with individual fields in space and time.

Key words: Climate, agricultural productivity, C4.5

alogarithm, prediction

I.INTRODUCTION

Crop production is a complex phenomenon that

is influenced by agro-climatic input parameters.

Agriculture input parameters varies from field to

field and farmer to farmer. Collecting such

information on a larger area is a daunting task.

However, the climatic information collected in India

at every 1sq.m area in different parts of the district

are tabulated by Indian Meteorological Department.

The huge such data sets can be used for predicting

their influence on major crops of that particular

district or place. There are different forecasting

methodologies developed and evaluated by the

researchers all over the world in the field of

agriculture or associated sciences. Some of the such

studies are : Agricultural researchers in Pakistan

have shown that attempts of crop yield maximization

through pro-pesticide state policies have led to a

dangerously high pesticide usage. These studies

have reported negative correlation between pesticide

usage and crop yield[1]. In their study they have

shown that how data mining integrated agricultural

data including pest scouting, pesticide usage and

meteorological data are useful for optimization of

pesticide usage. Thematic information related to

agriculture which has spatial attributes was reported

in one of the study[6]. Their study aimed at

discerning trends in agriculture production with

references to the availability of inputs. K- means

method was used to perform forecasts of the

pollution in the atmosphere [4], the k nearest

neighbor was applied for simulating daily

precipitations and other weather variables [11], and

different possible changes of the weather scenarios

are analyzed using SVMs[13]. Data mining

techniques are often used to study soil

characteristics. As an example, the k-means

approach is used for classifying soils in combination

with GPS-based technologies [14]. Apples were

checked using different approaches before sending

them to the market.[9], uses a k-means approach to

analyze color images of fruits as they run on

conveyor belts. [12] uses X-ray images of apples to

monitor the presence of water cores, and a neural

network is trained for discriminating between good

and bad apples. Spatial data mining introduced

especially decision tree algorithm applying to

agriculture land grading[15]. He combined spatial

data mining techniques with expert system

techniques and applied them to establish an

intelligent agriculture land grading information

system. The author adopted decision tree C4.5

algorithm and implement with Mo2.0 and VC++6.0

to build agriculture land grading expert system. The

study showed the advantages of this method in

addressing problems in land grading. A decision

tree classifier for agriculture data was proposed

[5].This new classifier uses new data expression and

can deal with both complete data and in complete

data. In the experiment,10-fold cross validat

**BEST CROP PREDICTION**

**BY**

**CALLBACK CATS**

**A.SUSMITHA**

**S.SWATHI**

**B.VYSHNAVI**

**B.LAVANYA**

**ABSTRACT**

The project is based on to predict the data regarding farming which is useful to predict the crop which is suitable to that land by basing on the previous datasets and the profit rates. It is providing the predictions the profit rate is increased which is very useful to farmers. Our constitution and the government main aim is also to provide good facilities to farmers and increase cultivation for the future society. There are so many techniques that are used to predict the moisture and so n for the particular land but it can’t tell about the perfect results that are going to face. Because it depends on climatic conditions are marketing rates. By using this project we are predicting the suitable crop for that particular land and also the marketing price regarding the previous data base. The prediction is based on previous production data of crops i.e, identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner. It is user friendly project where a normal person also able to use with little instructions and is helpful to the future generation for choosing a particular crop for cultivating.

**INTRODUCTION**

Agriculture is one of the most important occupation practiced in our country. It is the broadest economic sector and plays an important role in overall development of the country. About 60 % of the land in the country is used for agriculture in order to suffice the needs of 1.2 billion people. Thus, modernization of agriculture is very important and thus will lead the farmers of our country towards profit.

However, as the conditions change day by day very rapidly, farmers are forced to cultivate more and more crops. Being this as the current situation, many of them don’t have enough knowledge about the new crops and are not completely aware of the benefits they get while farming them. Also, the farm productivity can be increased by understanding and forecasting crop performance in a variety of environmental conditions. The processing part also take into consideration two more datasets i.e. one obtained from weather department, forecasting the weather expected in current year and the other data being static data. This static data is the crop production and data related to demands of various crops obtained from various government websites. The proposed system applies machine learning and prediction algorithm like Classification to identify the pattern among data and then process it as per input conditions. This in turn will propose the best feasible crops according to given environmental conditions.

**Support vector Machine**

SVM classifier treated as one of the dominant classification algorithms. In further sections of our article, we were going to discuss linear and non-linear classes. However, SVM is a supervised learning technique. When we have a dataset with features & class labels both then we can use Support Vector Machine. But if in our dataset do not have class labels or outputs of our feature set then it is considered as an **unsupervised learning algorithm**. In that case, we can use Support Vector Clustering.

For a dataset consisting of features set and labels set, an SVM classifier builds a model to predict classes for new examples. It assigns new example/data points to one of the classes. If there are only 2 classes then it can be called as a Binary SVM Classifier.

There are 2 kinds of SVM classifiers:

1. **Linear SVM Classifier**
2. **Non-Linear SVM Classifier**

**SVM Linear Classifier**

In the linear classifier model, we assumed that training examples plotted in space. These data points are expected to be separated by an apparent gap. It predicts a straight hyperplane dividing 2 classes. The primary focus while drawing the hyperplane is on maximizing the distance from hyperplane to the nearest data point of either class. The drawn hyperplane called as a maximum-margin hyperplane.

**SVM Non-Linear Classifier**

 In the real world, our dataset is generally dispersed up to some extent. To solve this problem separation of data into different classes on the basis of a straight linear hyperplane can’t be considered a good choice. For this Vapnik suggested creating Non-Linear Classifiers by applying the kernel trick to maximum-margin hyperplanes. In Non-Linear SVM Classification, data points plotted in a higher dimensional space.

**OBJECTIVES**

The objectives of this study may therefore be outlined as follows.

1. To obtain estimates of aggregate physical production functions for the yields of various crops in specified states, considering various technological factors and a newly developed meterological weather index as inputs.

2. To obtain stochastic yield projections through 1971, based upon the estimated production functions and projected inputs, with weather as a stochastic input.

3. To derive simple decision models to demonstrate the usefulness of the stochastic yield projections in meeting specified agricultural policy goals. This study is a pilot study in that the crops and states considered are limited to com in Illinois and Iowa, grain sorghum in Kansas and Nebraska and wheat in Kansas, Nebraska and North Dakota, Time-series data on past yields will be regressed on observed weather data, fertilizer application rates and other technological inputs to provide estimates of yield production functions. Estimates of the errors associated with weather predictions will provide a basis for approximating probable errors associated with yield predictions as 3 derived from these production functions. Should the results of the present study warrant more comprehensive studies of a similar nature, the procedures developed and the results obtained will provide useful guidelines for such future analyses.

**Problem statement**

Remote Sensing data provides high quality spatial and temporal information about land surface features which include the environmental impacts on crop growth conditions. It has been proved to be an effective tool to assess and monitor vegetation parameters, crop vigour and yield estimation. However, the problem lies in the fact that most of the studies are conducted at a national/regional level covering very large areas. The use of low resolution images have resulted into generalisation of the crop condition and yield estimates. The coarse resolution also had a mixture of crops and other non-crop vegetation that was later correlated to the final crop yield. On the other hand, relatively few studies have been conducted based on the relationship between remote sensed data and field scale crop yield. At this particular stage, the yield or agricultural production is a result of several complex factors that use other external parameters to compute the yield. Agricultural production is also influenced by the following variabilities: yield variability, field variability, soil variability, crop variability, anomalous factor variability and management variability (Oliver et al., 2013; Zhang et al., 2002). Those variabilities result in differences in crop growth within agricultural fields that can be quantified by monitoring crop canopy variables throughout the growing season. Important variables in this context include leaf area index (LAI), biomass, and nitrogen status (Hansen and Schjoerring,2003; Serrano et al.,2000). Also, the crop insurance claim is usually calculated on the basis of crop cutting experiments. Therefore there has always been a problem in getting timely and accurate data, due to which payment of claims to farmers were getting delayed. Currently claim adjusters often have to physically walk out into a field to measure the extent of crop damage. Therefore, developing an assessment tool with the aid of UAV acquired imageries that will not only compute the yield potential of a particular crop but also estimate the amount of crop insurance that can be paid to farmers.

**INDUSTRIAL PROFILE**



**REVIEW OF LITERATURE**

Classification problems and their corresponding solving approaches constitute one of the fields of machine learning. The application of classification schemes in Renewable Energy (RE) has gained significant attention in the last few years, contributing to the deployment, management and optimization of RE systems. The main objective of this paper is to review the most important classification algorithms applied to RE problems, including both classical and novel algorithms. The paper also provides a comprehensive literature review and discussion on different classification techniques in specific RE problems, including wind speed/power prediction, fault diagnosis in RE systems, power quality disturbance classification and other applications in alternative RE systems. In this way, the paper describes classification techniques and metrics applied to RE problems, thus being useful both for researchers dealing with this kind of problem and for practitioners of the field. In the last decade, global energy demand has increased to non-previously seen levels, mainly due to the increase in population, fierce urbanization in developed countries and aggressive industrial development all around the world

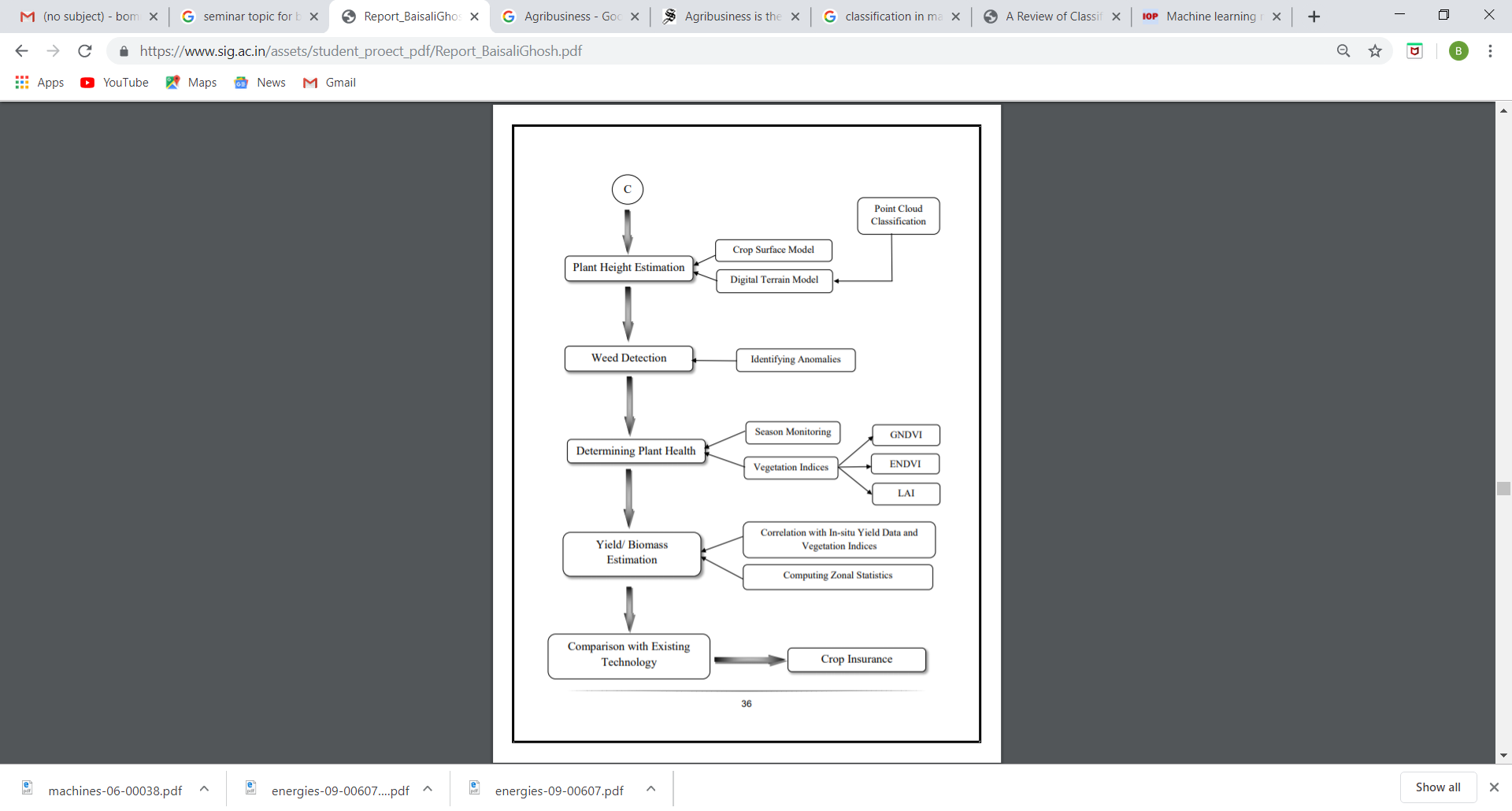
[1]. Conventional fossil-based energy sources have limited reservoirs and a deep environmental impact (contributing to global warming), and therefore, they cannot satisfy this global demand for energy in a sustainable way

[2]. These issues related to fossil-based sources have led to a very important development of Renewable Energy (RE) sources in the last few years, mainly in renewable technologies, such as wind, solar, hydro or marine energies, among others. The main problem with RE resources is their dependency on environmental conditions in the majority of cases (namely wind speed, solar irradiance or wave height) and the fact that individual renewable sources cannot provide continuous power supply because of their uncertainty and intermittent nature. A huge amount of research is being conducted to obtain a higher penetration of renewable resources into the electric system. The development of new and modern electric networks, including microgrids with renewable distributed generation, is, without a doubt, one of the main current research tracks in this topic, with a large amount of engineering sub-problems involved (such as microgrid topology design and operation optimization, microgrid control, optimal RE.

**METHODOLOGY**

A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes.  
For example, when filtering emails “spam” or “not spam”, when looking at transaction data, “fraudulent”, or “authorized”. In short Classification either predicts categorical class labels or classifies data (construct a model) based on the training set and the values (class labels) in classifying attributes and uses it in classifying new data. There are a number of classification models. Classification models include logistic regression, decision tree, random forest, gradient-boosted tree, multilayer perceptron, one-vs-rest, and Naive Bayes.

The methodology framed show the several aspects which can be used for yield estimation without extensive field work. The above methodology is followed to determine the health of the crop using remote sensing techniques on an aerial platform. This demonstrates how this technology can be simpler and time-saving as compared to the conventional techniques used for yield estimation and subsequent determination of crop insurance for farmers.



**EXPLORATORY DATA ANALYSIS**

Some of the most popular **tasks *T* in machine learning** are the following:

* classification of an instance to one of the categories based on its features;
* regression — prediction of a numerical target feature based on other features of an instance;
* clustering — identifying partitions of instances based on the features of these instances so that the members within the groups are more similar to each other than those in the other groups;
* anomaly detection — search for instances that are “greatly dissimilar” to the rest of the sample or to some group of instances;
* and so many more.

A good overview is provided in the “Machine Learning basics” chapter of [“Deep Learning”](http://www.deeplearningbook.org/) (by Ian Goodfellow, Yoshua Bengio, Aaron Courville, 2016).

**Experience *E*** refers to data (we can’t go anywhere without it). Machine learning algorithms can be divided into those that are trained in *supervised* or *unsupervised* manner. In unsupervised learning tasks, one has a *set* consisting of *instances* described by a set of *features*. In supervised learning problems, there’s also a *target variable*, which is what we would like to be able to predict, known for each instance in a *training set*.

**Example**

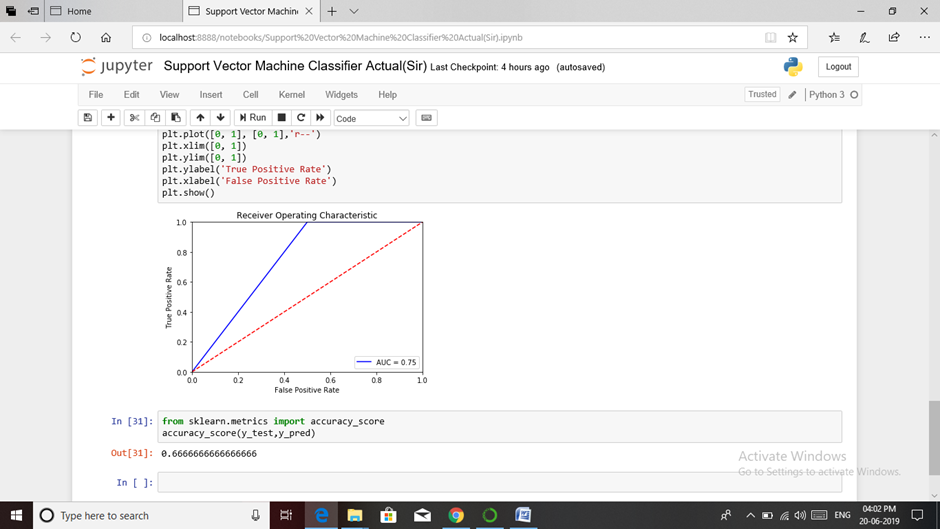
Classification and regression are supervised learning problems. For example, as a credit institution, we may want to predict loan defaults based on the data accumulated about our clients. Here, the experience *E* is the available training data: a set of *instances* (clients), a collection of *features* (such as age, salary, type of loan, past loan defaults, etc.) for each, and a *target variable* (whether they defaulted on the loan). This target variable is just a fact of loan default (1 or 0), so recall that this is a (binary) classification problem. If you were instead predicting *by how much time* the loan payment is overdue, this would become a regression problem.

Finally, the third term used in the definition of machine learning is a **metric of the algorithm’s performance evaluation *P*.** Such metrics differ for various problems and algorithms, and we’ll discuss them as we study new algorithms. For now, we’ll refer to a simple metric for classification algorithms, the proportion of correct answers — *accuracy* — on the test set.

**DATASET**

|  |  |  |  |
| --- | --- | --- | --- |
| Min temp | humidity | Rainfall | Crop |
| 16 | 85 | 150 | rice |
| 21 | 95 | 100 | wheat |
| 15 | 80 | 140 | rice |
| 20 | 95 | 150 | wheat |
| 23 | 70 | 200 | rice |
| 25 | 87 | 160 | wheat |
| 22 | 79 | 135 | rice |
| 24 | 93 | 145 | wheat |
| 19 | 99 | 210 | rice |
| 25 | 87 | 190 | wheat |
| 22 | 60 | 200 | rice |
| 24 | 93 | 145 | white |

**VISUALIZATION**



**CONCLUSION**

The proposed system takes into consideration the data related to soil, weather and past year production and suggests which are the best profitable crops which can be cultivated in theapropos environmental condition. As the system lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.

In the future, all farming devices can be connected over the internet using IOT. The sensors can be employed in farm which will collect the information about the current farm conditions and devices can increase the moisture, acidity, etc. accordingly. The vehicles used in farm like tractor will be connected to internet in future which will, in real time pass data to farmer about crop harvesting and the disease crops may be suffering from thus helping the farmer in taking appropriate action. Further the best profitable crop can also be found in light of the monetary and inflation ratio.

**REFERENCES**

* Aditya Shastry, H.A Sanjayand E.Bhanushree,“Prediction of crop using Regression Technique”,International Journal of computing12 (2):96-102 2017,ISSN:1816-9503
* E. Manjula , S. Djodiltachoumy,“A Model for Prediction of Crop”, International Journal of

Computational Intelligence and Informatics, Vol. 6: No. 4, March 2016

* Mrs.K.R.Sri Preethaa, S.Nishanthini, D.SanthiyaK.Vani Shree ,“Crop Prediction”,Internationa Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III,Issue III, March- 201
* Yethiraj N G , ” Applying data mining techniques in the field of Agriculture and allied sciences”, Vol01, Issue 02, December 2012