INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

CROP YIELD PREDICTION USING MACHINE LEARNING TECHNIQUES (SMART FARMING)

[1]Ch. KODANDARAMU, Associate Professor

[2]KANAPAREDDY JITENDRAKUMARI [3]KORADA RAMBABU

[4]KUBIREDDY SASIDHAR [5]GANAGALLA RAMESH

[1,2,3,4,5]Miracle Educational Society Group of Institutions, Vijayanagaram, Ap, India

ABSTRACT

As a coastal state, Tamil Nadu faces uncertainty in agriculture which decreases its production. Withmore population and area, more productivity should be achieved but it cannot be reached. Farmershave words-of-mouth in past decades but now it cannot be used due to climatic factors. Agriculturalfactors and parameters make the data to get insights about the Agri-facts. Growth of IT world drivessome highlights in Agriculture Sciences to help farmers with good agricultural information. Intelligenceof applying modern technological methods in the field of agriculture is desirable in this current scenario.Machine Learning Techniques develops a well-defined model with the data and helps us to attainpredictions. Agricultural issues like crop prediction, rotation, water requirement, fertilizer requirementand protection can be solved. Due to the variable climatic factors o f the environment, there is a necessity to have a efficient technique to facilitate the crop cultivation and to lend a hand to thefarmers in their production and management. This may help upcoming agriculturalists to have a betteragriculture. System o f recommendations can be provided to a farmer to help them in crop cultivation with the help o f data mining. To implement such an approach, crops are recommended based on itsclimatic factors and quantity. Data Analytics paves a way to evolve useful extraction from agriculturaldatabase. Crop Dataset has been analyzed and recommendation o f crops is done based on productivity

and season.

ISSN: 2057-5688

1. INTRODUCTION

Tamil Nadu being 7th largest area in India has 6th largest population. It is the leading producer of agricultureproducts. Agriculture is the main occupation of Tamil Nadu people. Agriculture has a sound tone in thiscompetitive world. Cauvery is the main source o f water. Cauvery delta regions are called as rice bow l ofTamilNadu. Rice is the major crop grown in Tamil Nadu. Other crops like Paddy, Sugarcane, Cotton, Coconut andgroundnut are grown. Bio-fertilizers are produced efficiently. Many areas Farming acts as major source ofoccupation.

Agriculture makes a dramatic impact in the economy o f a country. Due to the change of natural factors, Agriculture farming is degrading now-a-days. Agriculture directly depends on the environmental factors such assunlight humidity, soil type, rainfall, Maxim um and Minim um Temperature, climate, fertilizers, pesticides etc.

Knowledge of proper harvesting o f crops is in need to bloom in Agriculture. India has seasons ofWinter which occurs from December to M arch. Summer season from April to June. Monsoon or rainy season lasting from July to September and 4. Post-monsoon or autumn season occurringfrom October to November

Due to the diversity of season and rainfall, assessment of suitable crops to cultivate is necessary. Farmers facemajor problem s such as crop management, expected crop yield and productive yield from the crops. Farmers orcultivators need proper assistant regarding crop cultivation as now-a-days many fresh youngsters are interestedin agriculture.

Impact of IT sector in assessing real world problem is moving at a faster rate. Data is increasing day by day in field of agriculture. With the advancement in Internet o f Things, there are ways to grasp huge data in field ofAgriculture. There is a need o f a system to have obvious analyzes o f data o f agriculture and extract or use usefulinformation from the spreading data. To get insights from data, it has to be learnt.

ISSN: 2057-5688

2. LITERATURE SURVEY

[1] Shreya S. Bhanose, Kalyani A. Bogawar (2016) "Crop And Yield Prediction Model", InternationalJournal of Advance Scientific Research and Engineering Trends, Volume 1, Issue 1, April 2016

An agricultural sector necessitate for well defined and systematic approach for predicting the crops with its yieldand supporting farmers to take correct decisions to enhance quality of farming. The complexity of predicting thebest crops is high duet unavailability of crop knowledge-base. Crop prediction is an efficient approach for betterquality farming and increase revenue. Use of data clustering algorithm is an efficient approach in field of datamining to extract useful information and give prediction. Various approaches have been implemented so far areworked either for crop prediction.Crop prediction model aiding farmers to take correct decision. This indeedhelps in improving quality of farming and generate better revenue for farmers. Traditional clustering algorithmssuch as k-Means, improved rough k-Means and-means++ makes the tasks complicated due to random selection ofinitial cluster center and decision of number of clusters. Modified K-Means algorithm is thereby used to improvethe accuracy of a system as it achieves the high quality clusters duet initial cluster centric selection.

[2] Tripathy, A. K., et al.(2011) Data mining and wireless sensor network for agriculture pest/diseasepredictions.'s Information and Communication Technologies (WICT), 2011 World Congress on. IEEE.

Data driven precision agriculture aspects, particularly the pest/disease management, require a dynamic crop-weather data. An experiment was conducted in a semi-arid region to understand the crop-weather-pest/diseaserelations using wireless sensory and field-level surveillance data on closely related and interdependent pest(Thrips) - disease (Bud Necrosis) dynamics of groundnut crop. Data mining techniques were used to turn the datainto useful information/knowledge/relations/trends and correlation of crop-weather-pest/disease continuum.These dynamics obtained from the data mining techniques and trained through mathematical models werevalidated

ISSN: 2057-5688

with corresponding surveillance data. Results obtained from 2009 & 2010 kharif seasons (monsoon) and2009–10 & 2010–11 rabi seasons (post monsoon) data could be used to develop a real to near real-time decisionsupport system for pest/disease predictions

[3] Ramesh Babu Palepu (2017) " An Analysis of Agricultural Soils by using Data Mining Techniques",International Journal of Engineering Science and Computing, Volume 7 Issue No. 10 October.

Data mining is an approach through which in an synchronized manner we can find a workable solutionthat will be beneficial to increase the growth. The Farmers in agriculture sectors face a lot of issues anddifficulties due to the improper understanding and implementation of the activities to enhance theirgrowth and productivity. A large amount of data is available for analyses and scrutiny, however thoserelated to agriculture sector is in a small quantity. Hence segregation and processing of the same fromthe sources has to be done with proper methodology. Places having multiple grain growth and differentsoil structure makes it complex to have a perfect estimation of the crops yield both in quantity andquality. Creating a close link between the customer expectation and the producing capabilities of theagriculture sector can be win-win situation at both ends, this can be achieved with capturing datasegment wise and in a structured manner. Thus, the customer will be able to fulfil his requirement asper his wish, rather than being satisfied by what is being offered to him. The application of suchtechniques enables us to predict and make analysis of various problems and helps farmers to makedifficult farming decisions based on the conditions, soil fertility, crop duration, disease and otherimportant factors that can result in poor yield production

[4] Rajeswari and K. Arunesh (2016) "Analysing Soil Data using Data Mining ClassificationTechniques", Indian Journal of Science and Technology, Volume 9, May.

Soil is an essential key factor of agriculture. The objective of the work is to predict soil type using datamining classification techniques. Methods/Analysis: Soil type is predicted using data miningclassification

ISSN: 2057-5688

techniques such as JRip, J48 and Naive Bayes. These classifier algorithms are applied toextract the knowledge from soil data and two types of soil are considered such as Red and Black.Findings: In this paper, Data Mining and agricultural Data Mining are summarized. The JRip model canproduce more reliable results of this data and the Kappa Statistics in the forecast were increased.Application/Improvement: For solving the issues in Big Data, efficient methods can be created thatutilize Data Mining to enhance the exactness of classification of huge soil data sets.

[5] A.Swarupa Rani (2017), "The Impact of Data Analytics in Crop Management based on WeatherConditions", International Journal of Engineering Technology Science and Research, Volume 4,Issue5,May.

Many states faces uncertainty in agriculture which decreases its production. With more population andarea, more productivity should be achieved but it cannot be reached. Agricultural factors andparameters make the data to get insights about the Agri-facts. Growth of IT world drives some highlightsin Agriculture Sciences to help farmers with good agricultural information. The common difficultypresent among the Indian farmers are they don't opt for the proper crop based on their soil necessities.Because of this productivity is affected. This provides a farmer with sort of options of crops which will becultivated. Agricultural issues like crop prediction, rotation, water requirement, fertilizer requirementand protection can be solved. To implement such an approach, crops are recommended based on itsclimatic factors and quantity. Data Analytics paves a way to evolve useful extraction from agriculturaldatabase. Crop Dataset has been analyzed and recommendation of crops is done based on productivityand season.

[6] Pritam Bose, Nikola K. Kasabov (2016), "Spiking Neural Networks for Crop Yield Estimation Basedon Spatiotemporal Analysis of Image Time Series", IEEE Transactions On Geoscience And RemoteSensing.

This paper presents spiking neural networks (SNNs) for remote sensing spatiotemporal analysis of imagetime series, which make use of the highly parallel and low-power-consuming neuromorphic hardwareplatforms possible. This paper

ISSN: 2057-5688

illustrates this concept with the introduction of the first SNNcomputational model for crop yield estimation from normalized difference vegetation index image timeseries. It presents the development and testing of a methodological framework which utilizes the spatialaccumulation of time series of Moderate Resolution Imaging Spectroradiometer 250-m resolution dataand historical crop yield data to train an SNN to make timely prediction of crop yield. The research workalso includes an analysis on the optimum number of features needed to optimize the results from ourexperimental data set. The proposed approach was applied to estimate the winter wheat (Triticumaestivum L.) yield in Shandong province, one of the main winter-wheat-growing regions of China. Ourmethod was able to predict the yield around six weeks before harvest with a very high accuracy. Ourmethodology provided an average accuracy of 95.64%, with an average error of prediction of 0.236 t/haand correlation coefficient of 0.801 based on a nine-feature model.

[7] Priyanka P.Chandak (2017),” Smart Farming System Using Data Mining”, International Journal ofApplied Engineering Research, Volume 12, Number 11

Crop recommendation system or prediction system is the art of predicting crop yields to improve theproduction and production before the harvest actually takes place, it takes typically a couple of monthsin advance. Crop prediction depends on the computer programs that describe the plant-environmentand the soil features interactions in quantitative terms. The soil testing will start with the collections of asoil sample from the field. The first basic principles of the soil testing is that a field can be sampled insuch a way that by getting a chemical analysis of the soil sample and also majorly depend ontemperature and rainfall will accurately reflect the field&’ss true nutrient status on a particular area tohelp out farmers to improve the production.

3. EXISTING SYSTEM

Agriculture makes a dramatic impact in the economy o f a country. Due to the change of natural factors,Agriculture farming is degrading now-a-days. Agriculture directly depends on the environmental factors such

ISSN: 2057-5688

assunlight humidity, soil type, rainfall, Maxim um and Minim um Temperature, climate, fertilizers, pesticides etc.Knowledge of proper harvesting o f crops is in need to bloom in Agriculture. India has seasons ofWinter which occurs from December to March.Summer season from April to June. Monsoon or rainy season lasting from July to September and. Post-monsoon or autumn season occurring from October to November.

Due to the diversity of season and rainfall, assessment of suitable crops to cultivate is necessary. Farmers facemajor problem s such as crop management, expected crop yield and productive yield from the crops. Farers orcultivators need proper assistant regarding crop cultivation as now-a-days many fresh youngsters are interestedin agriculture.

3.1.LIMITATAION OF SYSTEM

The main challenge faced in agriculture sector is the lack of knowledge about the changing variations in climate.Each crop has its own suitable climatic features. This can be handled with the help of precise farming techniques.The precision farming not only maintains the productivity of crops but also increases the yield rate of production.The existing system which recommends crop yield is either hardware-based being costly to maintain, or not easilyaccessible. Despite many solutions that have been recently proposed, there are still open challenges in creating aan application with respect to crop recommendation

4. PROPOSED SYSTEM

Crop production depends on many agricultural parameters.Proposed work is based on the production of crops inprevious years, crops can be recommended to the farmers. This kind of suggestions will make farmer to knowthat whether that particular is yielding a good production in recent years. Production of crops may become lessdue to any crop disease, water problem and many other factors. While considering about the production, farmersmay get knowledge about which crop is in high volume in the market in that year. Based on this farmer can takedecision of trend on crops in recent years. Farmers will be given recommendation by considering the season ofcrop production. Tamilnadu Agriculture Dataset of about 1,20,000 records were taken. It contains fields like cropyear, crop

ISSN: 2057-5688

name, District, Season, Area cultivated and production. Recommendations were given to user based onthe production of crops, season when the crops cultivated

4.1 ADVANTAGES OF PROPOSED SYSTEM

The proposed model predicts the crop yield for the data sets of the given region. Integrating agriculture and MLwill contribute to more enhancements in the agriculture sector by increasing the yields and optimizing theresources involved. The data from previous years are the key elements in forecasting current performance. Theproposed system uses recommender system to suggest the right time for using fertilizers. The methods in theproposed system includes increasing the yield of crops, real-time analysis of crops, selecting efficient parameters,making smarter decisions and getting better yield.

5. MODULE DESCRIPTION

Many harvest yield expectation models have been developed. Bunching methods such as k-implies and k-means++ are used to collect data as groupings in order to predict agricultural yield [1]. Tripathy et al. [2] proposeda methodology for obtaining pesticide executives for crop development via an information mining process. Thenature of soil is a fundamental limit for agricultural investigation.In India, there are many different types of soil.Crops are produced based on the soil type in the area. The role of soil in advancing harvest development isdiscussed [3]. The dirt boundary is investigated using data mining techniques. The JRip, J48, and Naive Bayestechniques are used [4], resulting in more reliable results when dissecting red and black dirt. The impact ofagribusiness boundaries on crop executives is investigated in order to improve efficiency [5]. The farming factorsare being studied using brain organisations, delicate processing, large data, and fluffy rationale procedures.Pritam Bose [6] developed an SNN model for spatiotemporal analysis and crop evaluation.[7] A programmedframework was constructed to compile data regarding soil nature and weather patterns, using bunchingprocedures to separate the data and use it by ranchers in crop development. ICT-based communicationovercomes any barrier among farmers, such as language barriers. In

ISSN: 2057-5688

this day and age, mobile devicescommunicate information quickly. Ranchers can use Semantic Web-based Architecture [8] and GIS developmentsto learn about harvest ideas in a brief amount of time. GIS transmits data about climatic conditions andgeographic characteristics. Ranchers can then view this information using any ICT device. GIS and spatialdevelopments can reveal the universe&'ss monetary development [9]. Appropriate processes should be used toextract information from an enormous agriculture data source.Data Mining is an important aspect of theprocedures. By using mining, stowed useful information can be retrieved, as well as future forecasts. Theinformation gathered is organized;

5.1 Dataset Collection

The dataset comprising the soil specific attributes which are collected for Madurai district tested at soil testinglab, Madurai, Tamil Nadu, India. In addition, similar online sources of general crop data were also used. The cropsconsidered in our model include millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum,sugarcane, coriander. Figure 1 gives an analysis of the dataset. The number of instances of each crop available inthe training dataset is depicted. The attributes considered where Depth, Texture, Ph, Soil Color, Permeability,Drainage, Water holding and Erosion

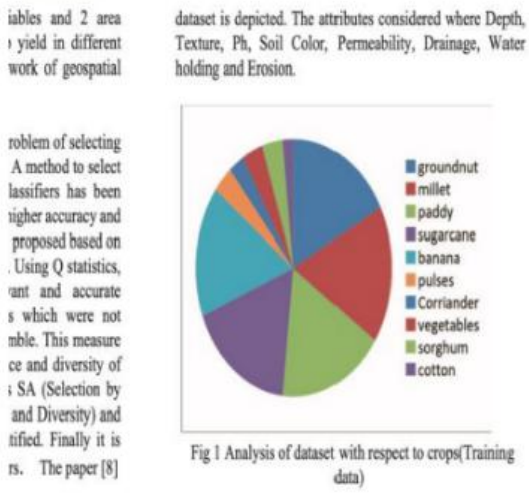


Fig 1 Analysis of dataset with respect to crops(Training data)

The above stated parameters of soil play a major role in the crop&'ss ability to extract water and nutrients from thesoil. For crop growth to their fullest potential, the soil must provide a satisfactory environment for it. Soil is theanchor of the roots. The water holding capacity determines the crop&'ss ability to absorb nutrients and othernutrients that are changed into ions, which is the form that the plant can use. Texture determines how porous thesoil is and the comfort of air and water movement which is essential to

ISSN: 2057-5688

prevent the plants from becomingwaterlogged. Soil texture which affects the soil&'ss ability to hold onto nutrients. The level of acidity or alkalinity(Ph) is a master variable which affects the availability of soil nutrients. The activity of microorganisms present inthe soil and also the level of exchangeable aluminum can be affected by PH. The water holding and drainagedetermine the penetration of roots. Hence for the following reasons the above stated parameters are consideredfor choosing a crop. Ensemble is a data mining model also known as the Committee Methods or ModelCombiners, that combine the power of multiple models to acquire greater prediction, efficiency than any of itsmodels could achieve alone. In our system, we use one of the most familiar ensembling technique called MajorityVoting technique .In the voting technique any number of base learners can be used. There has to be at least twobase learners. The learners are chosen in a way that they are competent to each other yet being complimentaryalso. Higher the competition higher is the chance of better prediction. But it is necessary for the learners to becomplimentary because when one or few members make an error, the probability of the remaining memberscorrecting this error would be high. Each learner builds itself into a model. The model gets trained using thetraining data set provided. When a new sample has to be classified, each model predicts the class on its own.Finally, the class which is predicted by majority of the learners is voted to be the class label of the new sample.This method is implemented in Rapid miner tool (figure 2, 3, 4, 5) depicts the process implemented in rapidminer.

	Crop_Year	Season	Crop	Area	Production
0	1997	Kharif	Banana	5619	183740.0
1	1997	Kharif	Horse-gram	6849	3040.0
2	1997	Kharif	Onion	2813	37188.0
3	1997	Kharif	Sesamum	1598	580.0
4	1997	Kharif	Small millets	63	50.0
...	...	...	...	...	...
537	2013	Whole Year	Sugarcane	1170	121181.0
538	2013	Whole Year	Sweet potato	2	42.0
539	2013	Whole Year	Tapioca	340	10174.0
540	2013	Whole Year	Tobacco	100	159.0
541	2013	Whole Year	Turmeric	1203	6472.0

5.2 Preprocess Dataset

It is a technique that is used to convert the raw data set into a clean data set.Prediction model creationWe create data into two models:A) Training modelB) Testing

ISSN: 2057-5688

modelThe division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percentrespectively.

5.3 Model evaluation

We apply the machine learning algorithm for testing part and get the accuracy of this model.PredictionThis module based on GUI part. we create a web page using bootstrap. The web page like (Nitrogen,Phosphorous, Potassium, PH value, Humidity, Rainfall, Temperature).now we get the data's from user to comparethe dataset values .finally it will predict for the Crop and soil to be planted.

6. INTERNAL MODULES

6.1 Numpy

numPy is a Python library used for working with arrays. It also has functions forworking in domain of linear algebra, fourier transform, and matrices. NumPy wascreated in 2005 by Travis Oliphant. It is an open source project and you can use itfreely. NumPy stands for Numerical Python.

NumPy arrays are stored at one continuous place in memory unlike lists, so processescan access and manipulate them very efficiently.This behavior is called locality of reference in computer science. This is the main reasonwhy NumPy is faster than lists. Also it is optimized to work with latest CPUarchitectures.

6.2 Pandas

Pandas is a Python library used for working with data sets. It has functions foranalyzing, cleaning, exploring, and manipulating data. The name 'sPandas's has areference to both 'sPanel Data's, and 'sPython Data Analysis's and was created by WesMcKinney in 2008.

Pandas allows us to analyze big data and make conclusions based on statisticaltheories. Pandas can clean messy data sets, and make them readable and relevant.Relevant data is very important in data science.

6.3 Matplotlib

Human minds are more adaptive for the visual representation of data rather thantextual data. We can easily understand things when they are visualized. It is better torepresent the data through the graph where we can analyze the data more efficientlyand make the specific decision according to data

ISSN: 2057-5688

analysis. Before learning thematplotlib, we need to understand data visualization and why data visualization isimportant.

Graphics provides an excellent approach for exploring the data, which is essentialfor presenting results. Data visualization is a new term. It expresses the idea thatinvolves more than just representing data in the graphical form (instead of usingtextual form).

This can be very helpful when discovering and getting to know a dataset and can helpwith classifying patterns, corrupt data, outliers, and much more. With a little domainknowledge, data visualizations can be used to express and demonstrate keyrelationships in plots and charts. The static does indeed focus on quantitativedescription and estimations of data. It provides an important set of tools for gaining aqualitative understanding.

6.4 Keras

Keras is an open-source high-level Neural Network library, which is written in Python iscapable enough to run on Theano, TensorFlow, or CNTK. It was developed by one ofthe Google engineers, Francois Chollet. It is made user-friendly, extensible, andmodular for facilitating faster experimentation with deep neural networks. It not onlysupports Convolutional Networks and Recurrent Networks individually but also theircombination.

It cannot handle low-level computations, so it makes use of the Backend library toresolve it. The backend library act as a high-level API wrapper for the low-level API,which lets it run on TensorFlow, CNTK, or Theano.Initially, it had over 4800 contributors during its launch, which now has gone up to250,000 developers. It has a 2X growth ever since every year it has grown. Bigcompanies like Microsoft, Google, NVIDIA, and Amazon have actively contributed to thedevelopment of Keras. It has an amazing industry interaction, and it is used in thedevelopment of popular firms likes Netflix, Uber, Google, Expedia, etc.

Focus on user experience has always been a major part of Keras. Large adoptionin the industry. It is a multi backend and supports multi-platform, which helps all

ISSN: 2057-5688

theencoders come together for coding. Research community present for Keras worksamazingly with the production community. Easy to grasp all concepts. 1It supports fastprototyping. It seamlessly runs on CPU as well as GPU. It provides the freedom todesign any architecture, which then later is utilized as an API for the project. 1It isreally very simple to get started with. Easy production of models actually makes Kerasspecial.

6.5 Tensorflow

TensorFlow is a software library or framework, designed by the Google team toimplement machine learning and deep learning concepts in the easiest manner. Itcombines the computational algebra of optimization techniques for easy calculation ofmany mathematical expressions.

6.6 Scikit-learn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning inPython. It provides a selection of efficient tools for machine learning and statisticalmodeling including classification, regression, clustering and dimensionality reductionvia a consistence interface in Python.

7. ALOGORITHMS USED

7.1 Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important featureis the capability of capturing descriptive decision making knowledge from the supplied data.Decision tree can be generated from training sets. The procedure for such generation based on the Set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consistsof a leaf labeled with this classStep 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S hasone outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si hasoutcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build asubsidiary decision tree by invoking the same procedure recursively on the set Si.

7.2 Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set ofindependent

ISSN: 2057-5688

(explanatory) variables. The name logistic regression is used when the dependent variablehas only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usuallyreserved for the case when the dependent variable has three or more unique values, such as Married,Single, Divorced, or Widowed. Although the type of data used for the dependent variable is differentfrom that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-responsevariables. Many statisticians feel that logistic regression is more versatile and better suited for modelingmost situations than is discriminant analysis. This is because logistic regression does not assume that theindependent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numericand categorical independent variables. It reports on the regression equation as well as the goodness of fit,odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysisincluding diagnostic residual reports and plots. It can perform an independent variable subset selectionsearch, looking for the best regression model with the fewest independent variables. It providesconfidence intervals on predicted values and provides ROC curves to help determine the best cutoff pointfor classification. It allows you to validate your results by automatically classifying rows that are not usedduring the analysis.

7.3 Naïve Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: itassumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (orabsence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervisedlearning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight anexplanation based on the representation bias. The naive bayes classifier is a linear

ISSN: 2057-5688

classifier, as well aslinear discriminant analysis, logistic regression or linear SVM (support vector machine). The differencelies on the method of estimating the parameters of the classifier (the learning bias).While the Naive Bayes classifier is widely used in the research world, it is not widespread amongpractitioners which want to obtain usable results. On the one hand, the researchers found especially it isvery easy to program and implement it, its parameters are easy to estimate, learning is very fast even onvery large databases, its accuracy is reasonably good in comparison to the other approaches. On the otherhand, the final users do not obtain a model easy to interpret and deploy, they does not understand theinterest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier tounderstand, and its deployment is also made easier. In the first part of this tutorial, we present sometheoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset withTanagra. We compare the obtained results (the parameters of the model) to those obtained with otherlinear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. Wenote that the results are highly consistent. This largely explains the good performance of the method incomparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2,Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

7.4 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regressionand other tasks that operates by constructing a multitude of decision trees at training time. Forclassification tasks, the output of the random forest is the class selected by most trees. For regressiontasks, the mean or average prediction of the individual trees is returned. Random decision forests correctfor decision trees&'s habit of overfitting to their training set. Random forests generally outperform decisiontrees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affecttheir performance.The first algorithm for random decision forests was created in 1995 by Tin

ISSN: 2057-5688

KamHo[1] using the randomsubspace method, which, in Ho&'ss formulation, is a way to implement the 'sstochastic discrimination'sapproach to classification proposed by Eugene Kleinberg.An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered'sRandomForests's as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman&'ss'sbagging's idea and random selection of features, introduced first by Ho[1] and later independently byAmit and Geman[13] in order to construct a collection of decision trees with controlled variance.Random forests are frequently used as 'sblackbox's models in businesses, as they generate reasonablepredictions across a wide range of data while requiring little configuration.

7.5 SVM

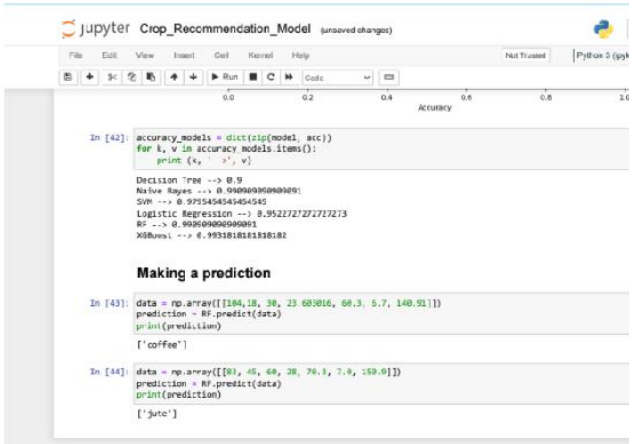
In classification tasks a discriminant machine learning technique aims at finding, based on anindependent and identically distributed (iid) training dataset, a discriminant function that cancorrectly predict labels for newly acquired instances. Unlike generative machine learningapproaches, which require computations of conditional probability distributions, a discriminantclassification function takes a data point x and assigns it to one of the different classes that are apart of the classification task. Less powerful than generative approaches, which are mostly usedwhen prediction involves outlier detection, discriminant approaches require fewer computationalresources and less training data, especially for a multidimensional feature space and when onlyposterior probabilities are needed. From a geometric perspective, learning a classifier is equivalentto finding the equation for a multidimensional surface that best separates the different classes inthe feature space.

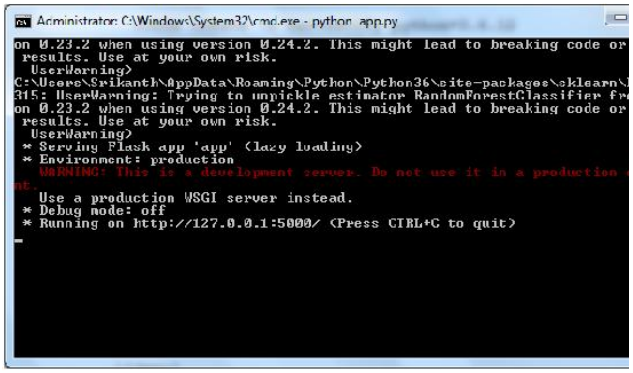
SVM is a discriminant technique, and, because it solves the convex optimization problemanalytically, it always returns the same optimal hyperplane parameter—in contrast to geneticalgorithms (GAs) or perceptrons, both of which are widely used for classification in machinelearning. For perceptrons, solutions are highly dependent on the initialization and terminationcriteria. For a specific kernel that transforms the data from the input space

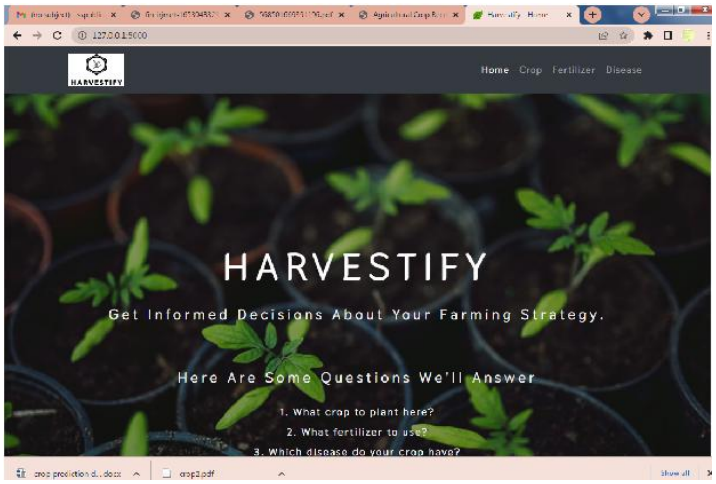
ISSN: 2057-5688

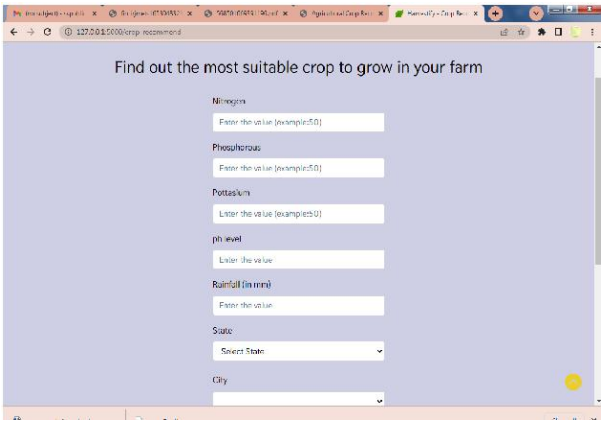
to the feature space,training returns uniquely defined SVM model parameters for a given training set, whereas theperceptron and GA classifier models are different each time training is initialized. The aim of Gasand perceptrons is only to minimize error during training, which will translate into severalhyperplanes' meeting this requirement.

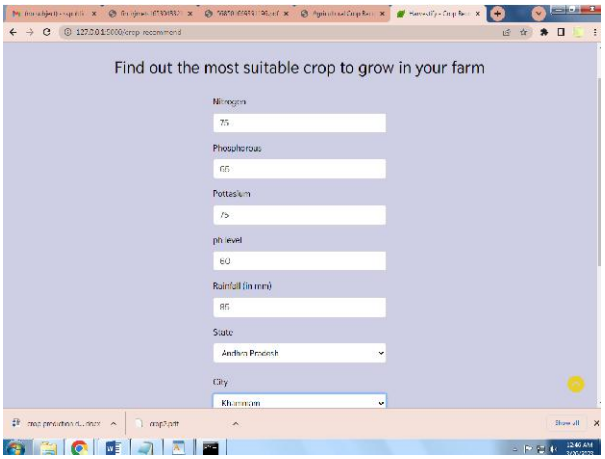
8. OUTPUT RESULTS





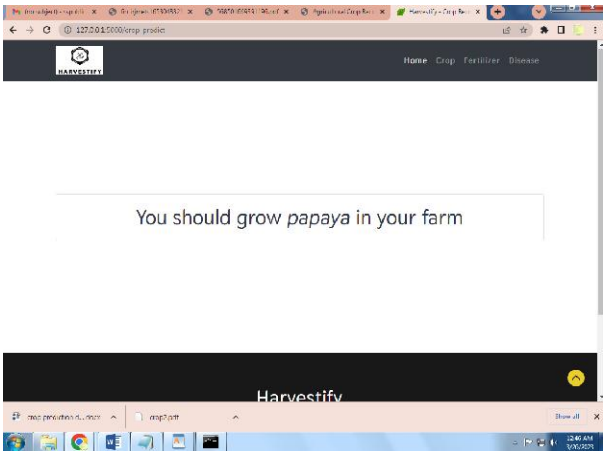


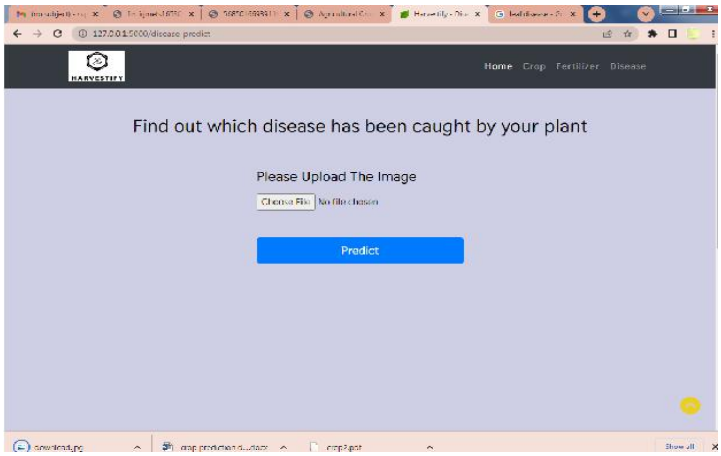


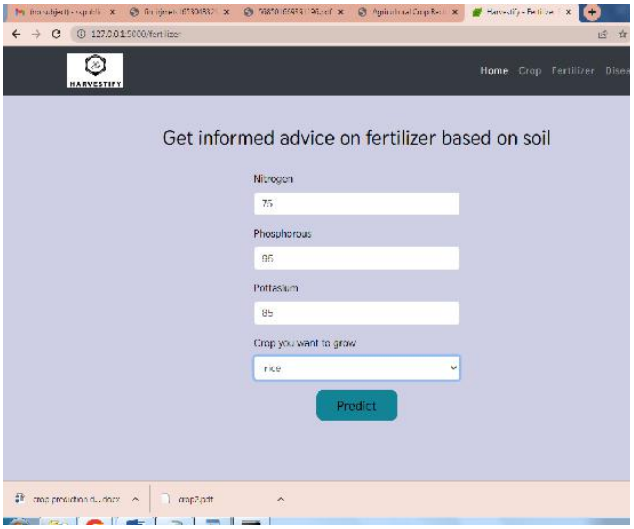


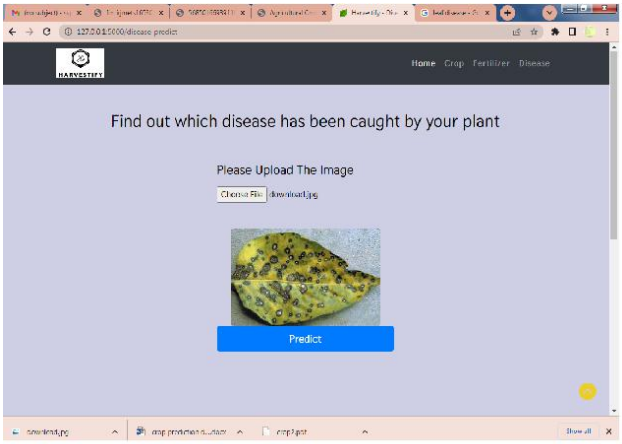
INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

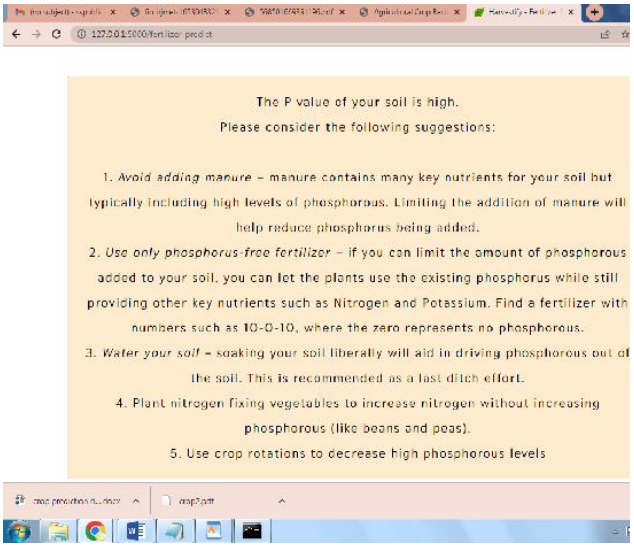
ISSN: 2057-5688











9. CONCLUSION

In this paper, significance of management of crops was studied vastly. Farmers need assistance with recenttechnology to grow their crops. Proper prediction of crops can be informed to agriculturists in time basis. ManyMachine Learning techniques have been used to analyze the agriculture parameters. Some of the techniques indifferent aspects of agriculture are studied

ISSN: 2057-5688

by a literature study. Blooming Neural networks, Soft computingtechniques plays significant part in providing recommendations. Considering the parameter like production andseason, more personalized and relevant recommendations can be given to farmers which makes them to yieldgood volume of production.

10. REFERENCES

[1] Shreya S. Bhanose, Kalyani A. Bogawar (2016) "Crop And Yield Prediction Model", International Journal ofAdvance Scientific Research and Engineering Trends, Volume 1,Issue 1, April 2016

[2] Tripathy, A. K., et al.(2011) 'sData mining and wireless sensor network for agriculture pest/diseasepredictions.'s Information and Communication Technologies (WICT), 2011 World Congress on. IEEE.

[3] Ramesh Babu Palepu (2017) " An Analysis of Agricultural Soils by using Data Mining Techniques", InternationalJournal of Engineering Science and Computing, Volume 7 Issue No. 10 October.

[4] Rajeswari and K. Arunesh (2016) "Analysing Soil Data using Data Mining Classification Techniques", IndianJournal of Science and Technology, Volume 9, May.

[5] A.Swarupa Rani (2017), "The Impact of Data Analytics in Crop Management based on Weather Conditions",International Journal of Engineering Technology Science and Research, Volume 4,Issue 5,May.

[6] Pritam Bose, Nikola K. Kasabov (2016), "Spiking Neural Networks for Crop Yield Estimation Based onSpatiotemporal Analysis of Image Time Series", IEEE Transactions On Geoscience And Remote Sensing.

[7] Priyanka P.Chandak (2017)," Smart Farming System Using Data Mining", International Journal of AppliedEngineering Research, Volume 12, Number 11.

[8] Vikas Kumar, Vishal Dave (2013), "KrishiMantra: Agricultural Recommendation System", Proceedings of the3rd ACM Symposium on Computing for Development, January.

[9] SavaeLatu (2009), "Sustainable Development : The Role Of Gis And Visualisation", The Electronic Journal onInformation Systems in Developing Countries, EJISDC 38, 5, 1-17.

ISSN: 2057-5688

[10] NasrinFathima.G (2014), "Agriculture Crop Pattern Using Data Mining Techniques", International Journal ofAdvanced Research in Computer Science and Software Engineering, Volume 4, May.

[11] Ramesh A.Medar (2014), "A Survey on Data Mining Techniques for Crop Yield Prediction", InternationalJournal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 9, September.

[12] ShakilAhamed.A.T.M, NavidTanzeem Mahmood (2015)," Applying data mining techniques to predict annualyield of major crops and recommend planting different crops in different districts in Bangladesh", ACIS 16thnternational Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/DistributedComputing (SNPD),IEEE,June.

[13] Shreya S.Bhanose (2016),"Crop and Yield Prediction Model", International Journal of Advence ScientificResearch and Engineering Trends, Volume 1,Isssue 1,ISSN(online) 2456- 0774,April.

[14] AgajiIorshase, OnyekeIdokoCharles,"A Well-Built Hybrid Recommender System for Agricultural Products inBenue State of Nigeria", Journal of Software Engineering and Applications,2015,8,581-589

[15] G. Adomavicius and A. Tuzhilin(2005), "Toward the Next Generation of Recommender Systems: A Survey ofthe State-of-theArt and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749,June.

[16] Avinash Jain, Kiran Kumar (2016),"Application of Recommendation Engines in Agriculture", InternationalJournal of Recent Trends in Engineering & Research, ISSN: 2455-1457.

[17] Kiran Shinde (2015),"Web Based Recommendation System for farmers", International Journal on Recent andInnovation Trends in Computing and Communication, Volume 3,Issue 3, ISSN:2321- 8169,March.

[18] Konstantinos G. Liakos, " Machine Learning in Agriculture: A Review",

ISSN: 2057-5688

Sensors 2018, 18, 2674;doi:10.3390/s18082674

[19] S.Vaishnavi, M.Shobana, N Geethanjali, Dr.S.Karthik, "Data Mining: Solving the Thirst of Recommendations toUsers", IOSR Journal of Computer Engineering (IOSR-JCE), Vol.16, no.6, 2014