AGRICULTURAL CROP PRODUCTIVITY AND RECOMMENDATIONS USING ML

Dr. I.Satyanarayana[1], P.sumana[2],Palle Jhansi [3], Naripaddi kavya [4] , Podila Sukanya [5], N.Hemanth Rao[6]
[1] Professor & Principal, Dept of ME Sri Indu Institute of Engineering & Technology, Hyderabad
[2] Assistant Professor, Dept of ECE Sri Indu Institute of Engineering & Technology, Hyderabad
[3-6] Student, Dept of ECE Sri Indu Institute of Engineering & Technology, Hyderabad.

***Abstract*— In today's time technology is playing a vital role in different sectors to overcome the difficulties and to have better and maximum results. In India, the farming sector has a huge impact on the Indian economy. Half of the country's population is still employed in the agriculture sector. The agriculture industry is largely influenced by the natural conditions of its surroundings and hence faa ces several challenges in actual farming practices. Agriculture practices in the country are largely primitive and technological change in the sector is slow. Effective technology can be used to increase the yield to reduce the maximum possible challenges in this field. Most of the m es it is observed that farmers tend to sow the crop according to its market value and possible financial profits rather than taking factors like soil conditsustainableiitd Nabil,ityinton he accounts. This may lead to undesirable results for farmers and the nature of sttheilo. In today's time, technologies like machine learning and deep learning can becogame-changersers in such fields if they are uproperlynner. This paper will represent an effective use of such technology to provide maximum assistance to farmers in the area of crop recommendation.**

I. INTRODUCTION

In India farming is not considered as a iness but also has a huge impact on the social life of people which arwhoociated towith There are many festivals and social gatherings are celebrated in accordance with the different seasons and practices involved in farming. Hence larga e part of population is dependent on agriculture field directly or indirectly. But the situation of farmers is not good in India. Despite employing half the population in agriculture, the agriculture sector contributes to only 20% of Indian GDP. Hence, it is in dire need of improvement in order to make a good and profitable yield and also, a practical need without harming the nature. That's where technology comes in and can have major effects on agricultural sector. Basically, our project aims to tackle the difficulties faced by the farmers and aims to provide a correct crop for the farmers to grow and avoid the undesirable results by providing effective solutions using machine learning techniques.

Given the population growth in the last decade, it is becoming extremely difficult to ignore the importance of agriculture, that has developed a lot in India. "site-specific" farming is the key to precision agriculture although precision agriculture has achieved better enhancement it is still facing certain issues. Precision agriculture plays an importatant role in the recommendation of crops. The recommendation of crops is dependent on various parameters.

Precision agriculture focuses on identifying these parameters in a site-specific way to identity issues. Not all the result given by precision agriculture are accurate to result but in agriculture, it is significant to have accurate and errors it may lead to heavy material and capital loss. Precision agriculture plays an importatant role in the recommendation of crops. The recommendation of crops is dependent on various parameters. Not all the result given by precision agriculture are accurate to result but in agriculture, it is significant to have accurate and errors it may lead to heavy material and capital loss.

II. PROPOSED SYSTEM

In our framework, we have proposed a procedure that is sepearated into various stages as appeared in above fig

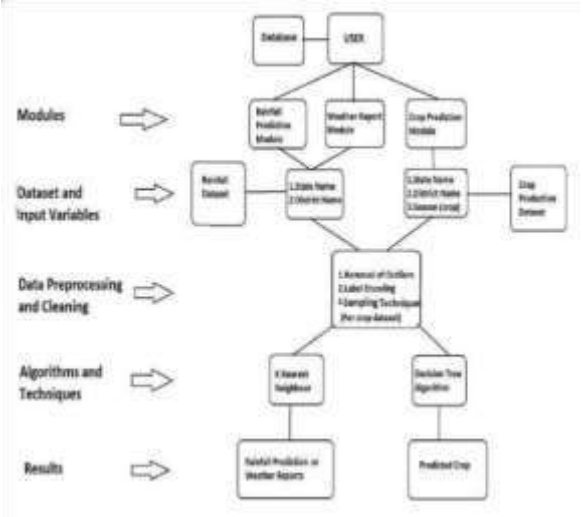


Fig: Block Diagram of Overall Methodology of proposed System

The phases are as per the following:

1. Collection of Datasets
2. Pre-processing
3. Feature Extraction
4. Applied various Machine Learning Algorithm

Dataset must have following attributes

1. Soil Parameters:
Soil Type
Soil ph value 2.
Climatic Parameters:
Humidity
Temperature
Wind
Rainfall
3. Production
Cost of cultivation
Previous year yield details for that region.

In this project we are performing crops prediction for district level. So main aim is to find the dataset which contains production details of past 10-12 years also details about climatic parameters and soil parameters like rainfall, temperature, moisture, soil contents etc. details. These factors will help in the prediction of the crops by using various classifiers on the given dataset. Thus, various factors are assessed and the factors strongly leading to accurate prediction of the crops.

The dataset that is used needs to be pre-processed because of the presence of redundant attributes, noisy data in it. Initially, data cleaning operation is performed where the redundant factors are determined and are not considered for the prediction of crops. Over18 which are either having the same values for all the employees or are completely unrelated to the prediction task. As part of the exploratory data analysis, the categorical factors are split and are assigned values as 0 and 1 based on whether the factor is present or not. These assigned values assist in further classification.

Data Collection:

The dataset consists of parameters like Nitrogen, Phosphorous, Pottasium, PH value of soil, Humidity, Temperature and Rainfall. The dataset have been obatained from the kaggle website. The dataset has 2200 istance or data that value taken from the past historic data. This dataset include eleven different crops such as rice, maize, chickpea, kidneybeans, oigeopeans, mothbeans, mungbeans, apple, orange,papaya,coconut, cotton,jute,and coffee.

Pre-Processing:

For the successful applications pre-processing is required. The data which is acquired from different resources are sometimes in raw form. It may contain some incomplete, redundant inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized.

The quality of input data can be improved through data pre-processing, also as data preparation, which in turn affects performance and analytical efficiency of the results. In this steps, the data was converted into the similar format to enhance features. The following processes were performed: The unit of temperature for some crops were in Fahrenheit and some crops were in Celsius; hence, everything was converted to the same unit. For experimentation purpose, Data augmentation process is carried out based on minimum and maximum values, each crop data row was duplicate into fifteen rows with 0.1 increases and decrease in their values for analysis. For missimg values in a data set, a The data which is acquired from different resources are sometimes in raw form. It may contain some incomplete, redundant inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized. The dataset that is used needs to be pre-processed because of the presence of redundant attributes, noisy data in it. Initially, data cleaning operation is performed where the redundant factors are determined and are not considered for the prediction of crops. Over18 which are either having the same values for all the employees or are completely unrelated to the prediction task.

III. LITERATURE REVIEW

The following research papers were referred by us before doing our project. While referring each of these papers we have come across various different findings discussed below.Each of the below entries discusses the title of the paper, the algorithms used and a general conclusion drawn from that research paper.

Title: Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques.
Author: Archana Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay.
Method Used: Random Forest, K- means clustering algorithm.
Remark: Accuracy of random forest is found to be higher than ID3(Iterative Dichotomiser 3) algorithm for crop prediction and K- means for fertilizer recommendation.

Title: Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine LearninAuthor: Keerthan Kumar T G, Shubha C, Sushma S A.
Method Used: Random Forest, Gaussian Naïve Bayes, Support Vector Machine, Linear Regression. Remark: In case of Crop Prediction, Random Forest proves to be a better classifier as compared to Gaussian Naïve Bayes and Support Vector Machine while linear regression works efficiently for grading soil.

Title: Soil Classification using Machine Learning Methods and Crop Suggestion Based on soil series.
Author: Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra. Method Used: Weighted K-NN, SVM, Bagged Tree.

Remark: SVM has given the highest accuracy in soil classification as compared to K-NN and Bagged tree algorithms.

Title: AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms.
Author: Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah. Method Used: Decision Tree, Random Forest, K- NN.
Remark: Accuracy rates were Decision Tree (90.20), KNN (89.78), Random Forest (90.43).

Title: Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques.
Author: Archana Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay.
Method Used: Random Forest, K- means clustering algorithm.
Remark: Accuracy of random forest is found to be higher than ID3(Iterative Dichotomiser 3) algorithm for crop prediction and K- means for fertilizer recommendation.

Title: Soil Classification using Machine Learning Methods and Crop Suggestion Based on soil series. Author: Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra.
Method Used: Weighted K-NN, SVM, Bagged Tree. Remark: SVM has given the highest accuracy in soil classification a compared to k

IV. SYSTEM REQUIREMENTS

Hardware Requirements

- Processor Type ` : Pentium i3
- Speed : 3.40GHZ
- RAM : 4GB DD2 RAM
- Hard disk : 500 GB
- Keyboard : 101/102 Standard Keys
- Mouse : Optical Mouse

Software Requirements

- Operating System : Windows 8
- Front end : JUPYTER NOTEBOOK
- Coding language : PYTHON

The system projected during this work is created by a neural network wherever inputs area unit treated on an individual basis. Static soil information in handled by fullyconnected layers whereas dynamic meteorological information is handled by continual LSTM layers. This explicit design was trained with historical information for many soil properties, precipitation, minimum and most temperature against historical yield labels at county level. When training, the model was tested in an exceedingly separate information set and showed comparable results with existing yield prognostication ways that create use of indepth remote sensing data. the most important lesson learnt from our experiments is that it's attainable get ascendable yield forecast as a result of the projected neural network model will notice and exploit redundant info each within the soil and within the weather information. To boot, the model might be able to learn AN implicit illustration of the cycles of the crops evaluated during this paper, considering the seasonal atmospherically information used as input.

The present study provides the potential use of information mining techniques in predicting the crop yield supported the environmental condition input parameters. The developed webpage is user friendly and therefore the accuracy of predictions square measure higher than seventyfiveper cent all told the crops and districts designated within the study indicating higher accuracy of prediction. The userfriendly web content developed for predicting crop yield may be utilized by any user their alternative of crop by providing environmental condition knowledge of that place.

The planned work introduces efficient degree economical crop recommendation system. Use of naïve mathematician makes the model terribly economical in terms of computation. The system is scalable because it may be wont to take a look at on totally different crops. From the yield graphs the simplest time of sowing, plant growth and gather of plant may be known. Conjointly the best and worst condition may also be incurred. The model focuses on all style of farms, and smaller farmers may also be benefitted.

This model may be more increased to seek out the yield of each crop, and for chemical recommendation.
Conjointly it may be changed to recommend concerning the fertilizers and irrigation want of crops.

In this study, we've given the analysis potentialities for the classification of soil by mistreatment well-known classification algorithms as J48, BF Tree, and OneR and Naïve Bayes; in data processing. The experiment was conducted on information instances from Kasur district, Pakistan. We have ascertained the comparative analysis of those algorithms have the various level of accuracy to determine the effectiveness and potency of predictions. However, the advantages of the higher understanding of soils classes will improve the productivity in farming, reduce dependence on fertilizers and build higher prognostic rules for the advice of the rise in yield. In the future, we have a tendency to contrive to form a Soil Management.

Two supervised classification machine learning formula has been enforced during this study. the choice Tree LearningID3 (Iterative Dichotomiser 3) and KNNR discover the patterns within the knowledge set containing average temperature and precipitation worth obtained throughout the cropping amount of six major crops in 10 major cities of Bangladesh for the past twelve years and provides the prediction. ID3 uses the choice tree table that consists of the ranges of the precipitation, temperature and yield knowledge. The research provides an answer to the current downside that

was much required for farmers in People's Republic of Bangladesh. Though the research is restricted to some mounted dataset, the long run ahead promises addition of a lot of knowledge which will be analyzed with more machine learning techniques to come up with crop predictions with higher exactness. Moreover, the analysis will result in profits and invention of advanced farming techniques which will improve our economy and can facilitate United States stand out as a technologically advanced count.

This model may be more increased to seek out the yield of each crop, and for chemical recommendation. Conjointly it may be changed to recommend concerning the fertilizers and irrigation want of crops.

The present study provides the potential use of information mining techniques in predicting the crop yield supported the environmental condition input parameters. The developed webpage is user friendly and therefore the accuracy of predictions square measure higher than seventyfiveper cent all told the crops and districts designated within the study indicating higher accuracy of prediction.

The experiment was conducted on information instances from Kasur district, Pakistan. We have ascertained the comparative analysis of those algorithms have the various level of accuracy to determine the effectiveness and potency of predictions. However, the advantages of the higher understanding of soils classes will improve the productivity in farming, reduce dependence on fertilizers and build higher prognostic rules for the advice of the rise in yield. In the future, we have a tendency to contrive to form a Soil Management.

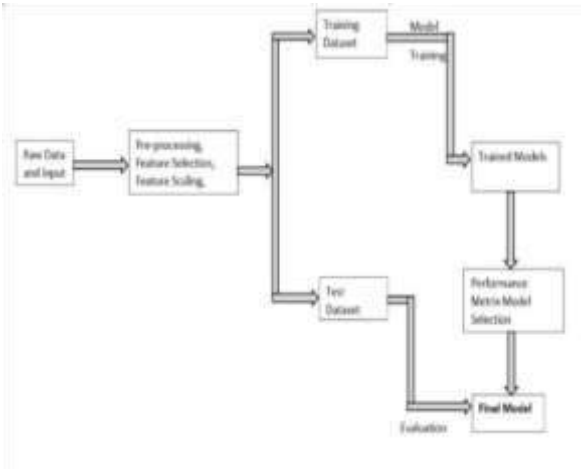
V. FEATURE SELECTION

It is important that we select only those features that will be necessary to determine the type of crop to grow. For this, we have created a correlation matrix that shows the linear relationship of a feature with every other feature. If features are highly correlated then that feature should be dropped, but as we can see in the below matrix that the features are not highly correlated with each other, hence it makes sense not to drop any of them and hence we will be
using all of them to predict the type of crop to grow

After performing the cleaning and preprocessing of the data, we perform data analysis and visualizations on our dataset. We try to analyze our data more clearly to find any trends or patterns in the dataset. We have created several visualizations of our dataset in order to understand the data properly. We have created bar charts, scatter plots, box plots etc. in order to visualize the data and find if there are any trends or patterns which we can find that will be useful while implementing our project.

One of the first steps is to make sure that the dataset we are using is accurate. The dataset should not have any missing values and if the dataset does have missing values, they should be replaced by the appropriate values. The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed. The skew value of the features should be checked and if the features have skewness, then those features should be normalized by using transformations. The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset.

VI. ULG DIAGRAM



The system projected during this work is created by a neural network wherever inputs area unit treated on an individual basis. Static soil information in handled by fully-connected layers whereas dynamic meteorological information is handled by continual LSTM layers. This explicit design was trained with historical information for many soil properties, precipitation, minimum and most temperature against historical yield labels at county level. When training, the model was tested in an exceedingly separate information set and showed comparable results with existing yield prognostication ways that create use of in-depth remote sensing data. the most important lesson learnt from our experiments is that it's attainable get ascendable yield forecast as a result of the projected neural network model will notice and exploit redundant info each within the soil and within the weather information. To boot, the model might be able to learn AN implicit illustration of the cycles of the crops evaluated during this paper, considering the seasonal atmospherically information used.

VII. IMPLEMENTATION

technology descriptionLibraries

Numpy: Numpy stands for numerical python. As the name gave it away, it’s an open source library for the Python programming language.

I hear you thinking: “another library…” but no such thing is true! Numpy is one of the most useful libraries especially if you’re crunching numbers Purpose of numpy:

Numpy adds support for large, multi-dimensional matrices and arrays, along with a gigantic collection of topend mathematical functions to operate on these arrays and matrices

It's objective is to make it easier for you to transform difficult functions or calculate some data analysis Numpy's biggest advantage is its fastness. It's so much faster than using the built-in Python's functions.
For example, it lets you simply calculate the mean and median of a dataframe with a plain line of code for each:

How to install by using Anaconda Prompt

First off you need to install Numpy but only if you're not using Anaconda. To do so

- Pip install numpy
- Conda install numpy

How to import numpy

- Import numpy
- Import numpy as np
- from numpy import *
- from numpy import add, subtract

Data Cleaning and processing:

One of the first steps is to make sure that the dataset we are using is accurate. The dataset should not have any missing values and if the dataset does have missing values, they should be replaced by the appropriate values. The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed. The skew value of the features should be checked and if the features have skewness, then those features should be normalized by using transformations. The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset.

The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed. The skew value of the features should be checked and if the features have skewness, then those features should be normalized by using transformations. The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset.

The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset. After performing the cleaning and preprocessing of the data, we perform data analysis and visualizations on our dataset.
Feature Selection:

It is important that we select only those features that will be necessary to determine the type of crop to grow. For this, we have created a correlation matrix that shows the linear relationship of a feature with every other feature. If features are highly correlated then that feature should be dropped, but as we can see in the below matrix that the features are not highly correlated with each other, hence it makes sense not to drop any of them and hence we will be
using all of them to predict the type of crop to grow

After performing the cleaning and preprocessing of the data, we perform data analysis and visualizations on our dataset. We try to analyze our data more clearly to find any trends or patterns in the dataset. We have created several visualizations of our dataset in order to understand the data properly. We have created bar charts, scatter plots, box plots etc. in order to visualize the data and find if there are any trends or patterns which we can find that will be useful while implementing our project.

One of the first steps is to make sure that the dataset we are using is accurate. The dataset should not have any missing values and if the dataset does have missing values, they should be replaced by the appropriate values. The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed. The skew value of the features should be checked and if the features have skewness, then those features should be normalized by using transformations. The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset.

VIII. SYSTEM TESTING

Framework Testing is a kind of programming testing that is performed on a total incorporated framework to assess the consistency of the framework with the relating necessities. Framework testing recognizes absconds inside both the coordinated units and the entire framework. The consequence of framework testing is the noticed conduct of a part of a framework when it is tried.

Framework Testing is fundamentally performed by a testing group that is autonomous of the improvement group that assists with testing the nature of the framework unbiased. The steps of testing are involved for the proposed system as follow.

The dataset containing the soil specific attributes which are collected from Polytest Laboratories soil testing lab, Pune, Maharashtra, India. In addition, similar sources of general crop data were also used from Marathwada University. The crops considered in our model include groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. The number of examples of each crop available in the training dataset is shown. The attributes considered where Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding and Erosion.

The above stated parameters of soil play a major role in the crop's ability to remove water and nutrients from the soil. For crop growth to their possible, the soil must provide

acceptable environment for it. Soil is the anchor of the roots. The water holding capacity determines the crop's ability to absorb nutrients and other nutrients that are changed into ions, which is the form that the plant can use. Texture determines how porous the soil is and the comfort of air and water movement which is essential to prevent the plants from becoming waterlogged.

The level of acidity or alkalinity (Ph) is a master variable which affects the availability of soil nutrients. The activity of microorganisms present in the soil and also the level of exchangeable aluminum can be affected by PH. The water holding and drainage determine the infiltration of roots. Hence for the following reasons the above stated parameters are considered for choosing a crop.

IX. CONCLUSION

In this project, the purpose to solve the problem of lack of efficient farming has been solved by designing and implementing a recommendation system, giving suggestions on the type of crops to be grown. We evaluated the performance of our system using various metrics and concluded that our neural network model (AgroSysNN) performs better than the traditional machine learning algorithms available. We achieved the accuracy of 99.7% and loss of 0.003 which is one of the best. It is evident that this system proves to be very effective and can be deployed to be used in real world scenarios. The future scope project lies with training the model with large amount of data making it more reliable and deploying the system into real world use cases.

The project work introduces an efficient crop recommendation system using classifier models. The system is scalable as it can be used to test on different crops. From the yield graphs the best time of sowing, plant growth and harvesting of plant can also be found out along with prediction for crops. Decision tree shows poor performance when dataset is having more variations but naïve bayes provides better result than decision tree for such datasets. The combination classification algorithm like naïve bayes and decision tree classifier are better performing than use of single classifier model.

X. FUTURE SCOPE

The system can be enhanced further to add following functionality: The main future works aim is to improved dataset with large number of attributes.We need to build a model, which can classify between healthy and disease, predict which disease is it. To build website and mobile app for easy to use. As concerning future score, when the farmers sow a particular crop, there might face some issues or diseasesin the crop before harvesting. In that case, they can upload the photographs of the crop and the soil report.To build website and mobile app for easy to use.

XI. REFERENCES

[1] Nidhi H Kulkarni, Dr G N Srinivasan, Dr B M Sagar, Dr N K Cauvery (2018) “Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique” - 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions.

[2] Tanmay Banavlikar, Aqsa Mahir, Mayuresh Budukh, Soham Dhodapkar (2018) “Crop Recommendation System Using Neural Network” - International Research Journal of Engineering and Technology (IRJET).

[3] S. Pudumalar, E. Ramanujam, R. Harine Rajashreeń, C. Kavyań, T. Kiruthikań, J. Nishań (2016) “Crop Recommendation System for Precision Agriculture” - IEEE Eighth International Conference on Advanced Computing (ICoAC).

[4] D. Diepeveen and L. Armstrong (2008) “Identifying key crop performance traits using data mining” - School of Computer and Information Science, EdithCowan University, Australia.

[5] S. Kanaga Suba Raja, R. Rishi, E. Sundaresan, V. Srijit (2017) “Demand based crop recommender system for farmers” - IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR).

[6] Zeel Doshi, Rashi Agrawal, Subhash Nadkarni, Prof. Neepa Shah (2018) “AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms” - Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).

[7] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare (2017) “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique” - International Research Journal of Engineering and Technology (IRJET).

[8] K. R. Akshatha, K. S. Shreedhara (2018) “Implementation of Machine Learning Algorithms for Crop Recommendation Using Precision Agriculture” - International Journal of Research in Engineering, Science andManagement.

[9] Maharashtra Agriculture website - http://krishi.maharashtra.gov.in/1001/Home.

[10] Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra” Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series” 2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.