Crop Yield Prediction using Machine Learning Techniques

J.Manikandan[1], Tummala Tharun Kumar reddy[2], Dogiparthy Saideep[3] , Mittapally Harshith[4]

[1]Asst Professor, Dept. Of CSE, CMR Institute of Technology, Hyderabad-501401

[2,3,4]Eight Semester, Dept. Of CSE, CMR Institute of Technology, Hyderabad-501401, India

---***---

ABSTRACT

Agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity, hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.

1.INTRODUCTION

The main goal of agricultural planning is to achieve maximum yield rate of crops by using limited number of land resources. Many machine learning algorithms can help in improving the production of crop yield rate. Whenever there is loss in unfavorable conditions, we can apply crop selecting method and reduce the losses. And it can be used to gain crop yield rate in favorable conditions. This maximizing of yield rate helps in improving countries economy. We have some of the factors that influence the crop yield rate. They are seed quality and crop selection. We need test the quality of the seeds before sowing. As we know that good quality of seeds helps in getting more yield rate. And selection of crops depends upon two things that is favorable and unfavorable conditions. This can also be improved by using hybridization methods. Many researches are carried out to improve agricultural planning. The goal is to get the maximum yield of crops. Many classification methods are also applied to get maximum yield of crops. Machine learning techniques can be used to improve the yield rate of crops. The method of crop selection is applied to improve crop production. The production of crops may depend on geographical conditions of the region like river ground, hill areas or the depth areas. Weather conditions like humidity, rainfall, temperature, cloud. Soil type may be clay, sandy, saline or peaty. Soil composition can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value or carbon and different methods of harvesting. Many parameters are used for different crops to do different predictions. These prediction models can be studied by using researches. These predictions are classified as two types. One is traditional statistic method and other is machine learning techniques. Traditional method helps in predicting single sample spaces. And machine learning methods helps in predicting multiple predictions. We need not to consider the structure of data models in traditional method whereas we need to consider the structure of data models in machine learning methods. Existing System: Hence our farmers should know all the new technologies of machine learning and other new techniques. These techniques help in getting maximum yield of crops. Many techniques of machine learning are applied on agriculture to improve yield rate of crops. The production of crops may depend on geographical conditions of the region like river ground, hill areas or the depth areas. Weather conditions like humidity, rainfall, temperature, cloud. Soil type may be clay, sandy, saline or peaty. Soil composition can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value or carbon and different methods of harvesting. Many parameters are used for different crops to do different predictions. Disadvantages of existing system are:- As the farmers are facing many problems in agricultural sector, we need to minimize their problems, as the farmers don't know to use Weka tool, Low Efficiency.

PROPOSED SYSTEM: we have considered only the Naive Bayes method and K-Nearest Neighbor method. Using these two methods we can predict which crops to

---

be selected for their land and season. This application we can do single testing by giving input as crop name, season selected and place selected. We can use any method among KNN or NB method. As soon you give the input you can select the method and mine the results. The results will tell you the yield rate of that crop. And we can do multiple testing by analyzing the datasets. In analyzing it allows you to select a whole file at once and get the accuracy. Advantages of proposed system are:-High Efficiency, reduction of problems can be done by implementing new techniques on agriculture, this application helps them to predict the yield, this will help the farmers which crop to be selected for their land or the region.

2. SYSTEM DESIGN

Data Flow Diagram: The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
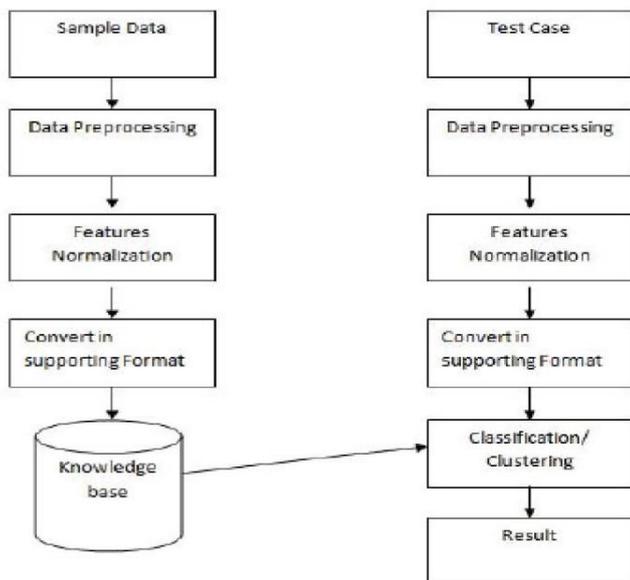

DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into

levels that represent increasing information flow and functional detail.
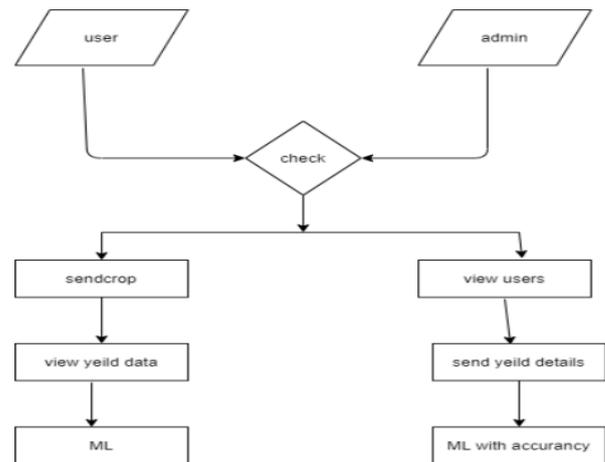
Use Case Diagram: A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.
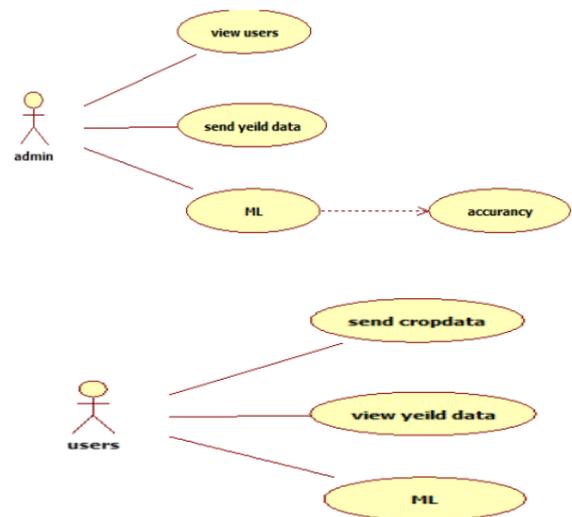

Class Diagram: In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.
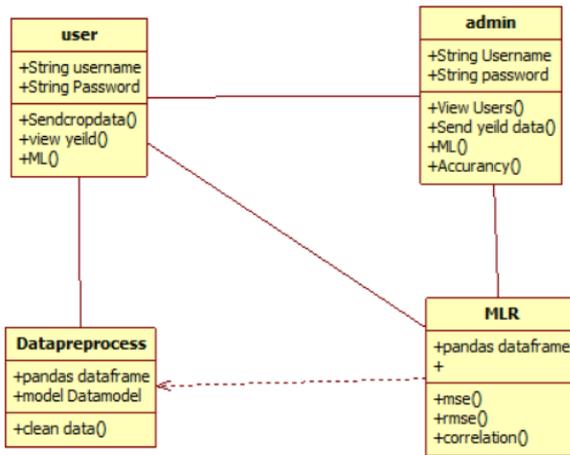
4.INPUT AND OUTPUT DESIGN

INPUT DESIGN:-The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things: What data should be given as input? How the data should be arranged or coded? The dialog to guide the operating personnel in providing input. Methods for preparing input validations and steps to follow when error occur.

Output Design:-A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements. Select methods for presenting information. Create document, report, or other formats that contain information produced by the system.

5. IMPLEMENTATIONS
5.1. PROJECT MODULES
➢ Server
➢ Remote User
➢ Data preprocess
➢ Feature Extraction
➢ Supervised Learning Models
➢ User Feedback Integration

MODULES DESCRIPTION:

5.1.1 Server: In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Upload Post, View All Uploaded Posts, View Rating Results, View Dislike Results, View Like Results, View Remote Users, View Posts Review, View Trending Posts, View All Recommended Posts. View and Authorize Users In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

5.1.2 Remote User: In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like SEARCH AND VIEW ALL POSTS, VIEW ALL POST REVIEWS, VIEW TRENDING POSTS, VIEW YOUR PROFILE, VIEW ALL POSTS RECOMMENDED.

5.1.3 Data Preprocess: Data preprocessing is a fundamental module that involves cleaning, transforming, and structuring raw email data. It includes tasks such as text normalization, removing special characters, and stemming. The module standardizes the input data, ensuring consistency and facilitating accurate feature extraction in subsequent stages.

5.1.4 Feature Extraction: Feature extraction involves identifying relevant attributes or features from preprocessed email content. Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings are applied to convert textual information into numerical vectors. These vectors capture semantic meaning, enabling the system to understand the context and nuances of email content **5.1.5 Supervised Learning Models:** Supervised learning models form the core of the spam detection system. Algorithms like Naive Bayes, Support Vector Machines (SVM), and Random Forest are trained using labeled data to recognize patterns in email features. These models learn to distinguish

between legitimate and spam emails based on the extracted features, enabling accurate classification during real-time email processing. 6.1.6 User Feedback Integration: The user feedback integration module allows users to provide feedback on identified emails. Positive feedback confirms accurate classifications, while negative feedback helps the system learn from misclassifications. User feedback is incorporated into the training data, enhancing the system's accuracy over time. Continuous learning and adaptation occur, ensuring the system remains effective against evolving spamming technique.

6. TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

6.1 TYPES OF TESTS Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. SYSTEM TEST System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points. White Box Testing White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level. Black Box Testing Black Box Testing is testing the

software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document.

6.2 TESTING LEVELS Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases. Integration Testing Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications. e.g. components in a software system or – one step up – software applications at the company level – interact without error. Acceptance Testing User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

7.RESULT

7.1 SCREEN SHOTS

7.1.1 Home page

//

7.1.2 User Registration:
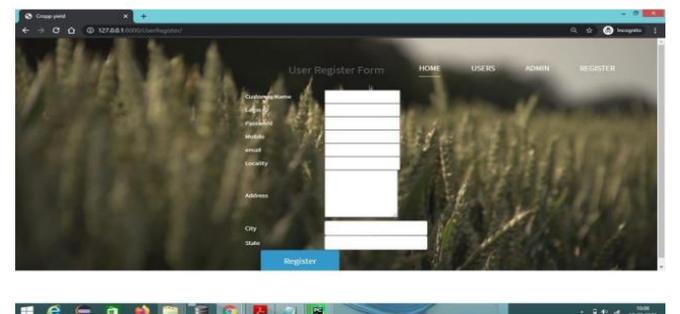
7.1.3 User login:
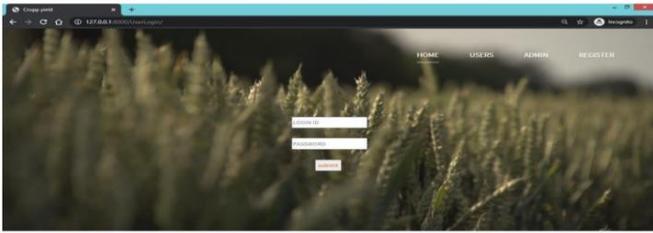
7.1.4 User Send Crops:
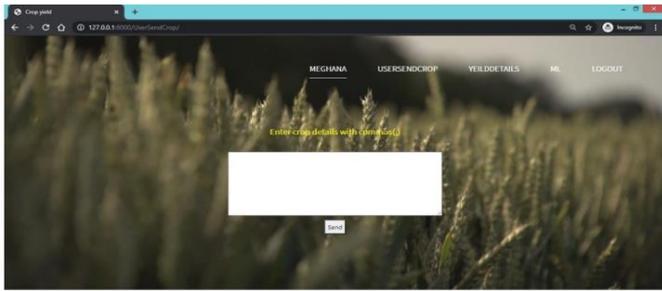
7.1.5 Yield data:
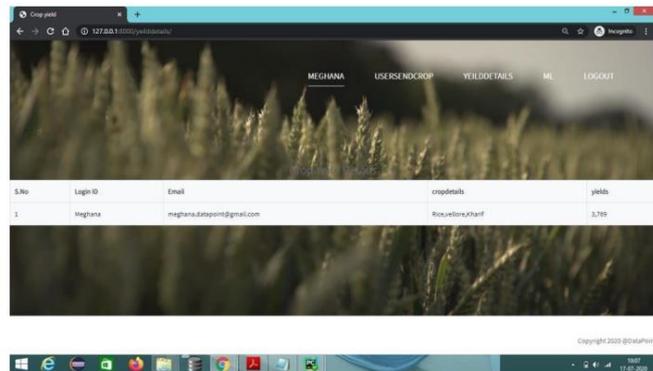
7.1.6 ML:
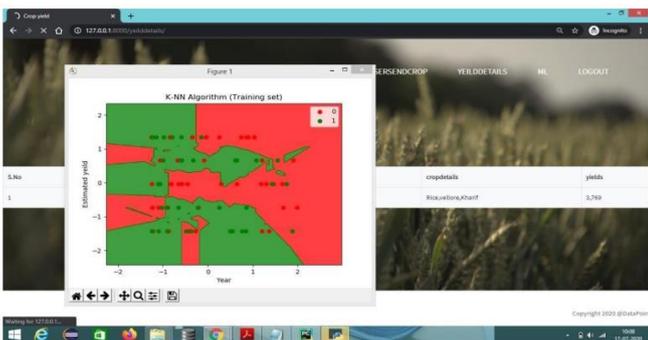
7.1.7 View Users:
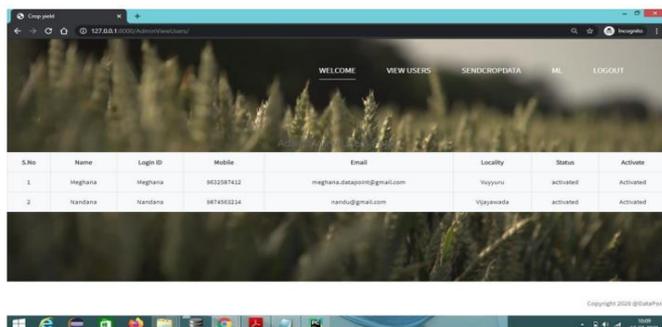
7.1.8 Send Crop data:
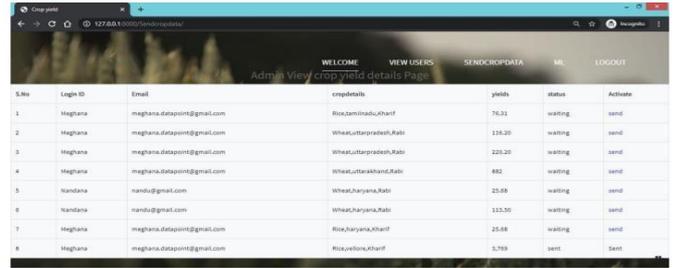
7.1.9 ML:
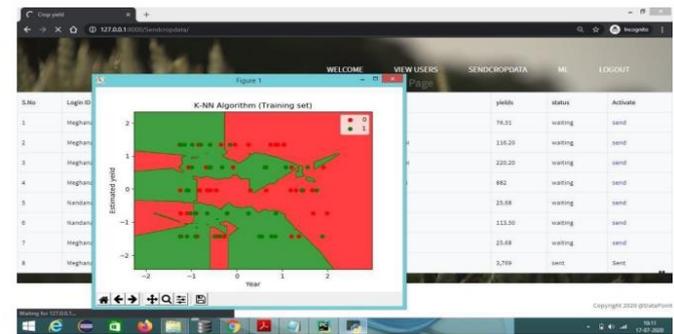
7.1.10 Accuracy:

8.CONCLUSIONS

Agriculture is the field which helps in economic growth of our country. But this is lacking behind in using new technologies of machine learning. Hence our farmers should know all the new technologies of machine learning and other new techniques. These techniques help in getting maximum yield of crops. Many techniques of machine learning are applied on agriculture to improve yield rate of crops. These techniques also help in solving problems of agriculture. We can also get the accuracy of yield by checking for different methods. Hence we can improve the performance by checking the accuracy between different crops. Sensor technologies are implemented in many farming sectors. This paper helps in getting maximum yield rate of the crops. Also helps in selecting proper crop for their selected land and selected season. These techniques will solve the problems of farmers in agriculture field. This will help in improving the economic growth of our country.

Future Enhancement:-Quantify recommendations for varieties: Currently the direction in which the factors

affect yield is available. I would like to quantify these effects and hopefully give more precise recommendations, incorporate more weather variables: Temperature, humidity, hours of sunlight, Incorporate soil data, Gather location-based variety data for recommendations: If all the varieties can be tested at the different locations, then varieties for which there wasn't data before can be recommended to particular locations, Attempt to use mixed models (treating some predictors as random variables).

9.REFERENCES

[1] J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique, International Journal on Engineering Technology, May 2015.

[2] A. Panangadan "Construction in a Simulated Environment using Temporal Goal Sequencing and Reinforcement Learning". Adaptive Behavior, 2009.

[3] Gour Hari Santra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016.

[4] Karan deep Kauri, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016.

[5] S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, March 2017.

[6] Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, Crop Selection Method Based on Various Environmental Factors Using Machine Learning, Feb -2017.

[7] D.Sindhura, B.Navya Krishna, K.Sai Prasanna Lakshmi,B.Mallikarjun Rao, Dr. J Rajendra Prasad, Effects of Climate Changes on Agriculture International Journal of Advanced Research in Computer Science and Software Engineering, March 2016.

[8] T.Giri Babu, Dr.G.Anjan Babu, Big Data Analytics to Produce Big Results in the Agricultural Sector, March 2016. . 36

[9] D Ramesh , B Vishnu Vardhan, Analysis Of Crop Yield Prediction Using Data Mining Techniques, International Journal of Research in Engineering and Technology, Jan-2015.7m

[10] Ashwani Kumar Kushwaha, SwetaBhattachrya, 5u

[11] Raorane A.A, Dr. Kulkarni R.V, Application Of Datamining Tool To Crop Management System, January 2015.

[12] Anshal Savla, Himtanaya Bhadada, Parul Dhawan, Vatsa Joshi, Application of Machine Learning Techniquesfor Yield Prediction on Delineated Zones in Precision Agriculture, May 2015.

[13] Siti Khairunniza-Bejo, Samihah Mustaffha and Wan Ishak Wan Ismail , Application of Artificial Neural Network in Predicting, Journal of Food Science and Engineering, January 20, 2014