Data-Driven Farming Techniques for Optimal Harvests

[1]Bodaballa Raghu Ram, [2]Chama Karthik Reddy

Student, Sathyabama University

Abstract - Indian economy is significantly influenced by the agricultural sector. The primary issue facing Indian farmers is that they frequently fail to select the appropriate crop for their soil. As a result, their productivity has suffered significantly. Precision agriculture has helped farmers solve their problems. Precision agriculture is a cutting-edge farming method that makes use of research data on soil types, characteristics, and crop production to recommend the best crop to farmers based on site-specific characteristics. This enhances production and decreases crop selection errors.In this project, we are developing an intelligent system that will help Indian farmers choose the best crop to cultivate based on the sowing season, the location of their farms, and the soil properties. Moreover, if the farmer plants the recommended crop, the system will predict the yield.

Keywords: Precision Agriculture, Yield PredictionMachine Learning, KNN Algorithm

1.INTRODUCTION

A farmer's decision about which crop to grow is generally influenced by his intuition and other irrelevant factors such as making quick profits, being unaware of market demand, overestimating a soil's ability to support a specific crop, and so on. A very poor decision on the farmer's part could put a significant strain on his family's financial situation. Perhaps this is one of the many factors contributingto the countless farmer suicides reported in the media on a daily basis. In a country like India, where agriculture and related sectors account for approximately 20.4 percent of GDP [2], such an erroneous judgment would have not only done this have a negative impact on the farmer's family, but it also has a negative impact on the entire economy of a region. As a result, we consider a farmer's dilemmaover which crop to grow during a specific season to be extremely serious. The need of the hour is to create a system that can provide Indian farmers with predictive insights, allowing them to make informed decisions about which crop to grow. With this in mind, we propose a system, an intelligent system, that takes into account environmental parameters (temperature, rainfall, geographical location in terms of state) and soil characteristics (pH value, soil type, and nutrient concentration) before recommending thebest crop to the user. Modern farms and agricultural operations are taken place more totally different than those many decades agone, primarily due to advancements in technology, as well as sensors, devices, machines, and knowledge technology. Agriculture today routinely employs sophisticated technologies such as robots, temperature and moisture sensors, aerial photography, GPS technology, and a plethora of complex IOT devices. These advanced devices in agriculture enable businesses and farmers to be additional profitable, efficient, safer, and more environmentally friendly therise of digital agriculture and its connected technologies has opened a wealth of latest knowledge opportunities. Remote sensors, cameras, and other connected devices will collect data 24 hours a day, seven days a week across an entire farm or land. These will monitor plant health, soil condition, temperature, humidity, etc. the quantity of information these sensors will generate is overwhelming. this enables farmers to achieve a far better an improved understanding of state of matters on the bottom through advanced technology which will inform them additional regarding their situation more accurately and quickly. The environmental datathat is gathered by remote sensors are processed by algorithms and statistical data which will be understood and helpful to farmers for decision makings and keep track of their farms. The more inputs and statistical data that are collected, the better the algorithmic rule predicts the outcomes. And the goal is for farmers to use these technologies to achieve their goal of improved harvest by making better field selections. The information captured is processed with an explicit algorithmic rule and passed to a centralized database that is connected to different modules of the research by implementing the system of temperature, soil hydrogen ion concentration, and soil wetness detection., so the main system will predict the most effective crop kind that the farmer should grow to require the most outcome of the crop kind that is farmed in a home garden or the respectable land area.

2.LITERATURE SURVEY

The paper [1] states the requirements and planning needed for developing a software model for precision farming is discussed. It deeply analyses the basics of precision farming. The author's start from the basics of precision farming and move towards developing a modelthat would support it. This paper describes a model that applies Precision Agriculture (PA) principles to small, open farms at the individual farmer and crop level, to affect a degree of control over variability. The comprehensive objective of the model is to deliver direct advisory services to even the smallest farmer at the levelof his/her smallest plot of crop, using the most accessible technologies such as SMS and email. This model has been designed for the scenario in Kerala State where the average holding size is much lower thanmost of India. Hence this model can be deployed elsewhere in India only with minor modifications. The paper [2] makes a comparative study of classification algorithms and their performance in yield prediction in precision agriculture. These algorithms are implementedin a data set collected for several years in yield prediction on soya bean crop. The algorithms used for yield prediction in this paper are Support Vector Machine, Random Forest Neural Network, REPTree, Bagging, and Bayes. The conclusion drawn at the end isthat bagging is the best algorithm for yield prediction among the above stated algorithms since the error deviation in bagging is minimum with a mean absolute error of 18985. 7864. The paper [3] states the necessityfor crop yield prediction and its help in a nation's strategic policy making in agriculture. It

facilitates flexible inclusion of various techniques towards crop yield prediction. A tool was also developed that would help people to predict crop yield for various crops with dependent and independent variables. The paper [4]states the usage of agricultural data with data miningand visual data mining techniques are depicted. This paper reduces the high dimensional agricultural data tosmaller size to acquire useful knowledge related to yield,input application. The techniques used is Self-organizingmaps and multi-dimensional scaling techniques toreduce the data. The conclusion derived is that Self-organizing maps is suitable when dataset is large issuitable when data set is small. The paper [5] depicts the importance of crop selection and the factors decidingthe crop selection like production rate, market price and government policies are discussed. This paper proposesa Crop Selection Method (CSM) which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crop to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence there exists a need to include a prediction method with improved accuracy and performance. Data mining techniques in paper [6] are used to estimate the crop yield for cereal crops in major districts of Bangladesh. The methodology comprises of two parts namely Clustering, Linear Regression, (ANN) artificial neural network in rapid miner tool. The accuracy of prediction lies in the range of 90-95. The data set included 5 environmental variables, 3 biotic variables and 2 area related variables to determine the crop yield in different districts. The paper proposed a future work of geospatialanalysis to improve accuracy. The paper [7] aims to solve the crucial problem of selecting the classifiers for the ensemble learning. A method to select a best classifier set from a pool of classifiers has been proposed. The proposal aims to achieve higher accuracy and performance. A method called SAD was proposed based on accuracy and classification performance. Using Q statistics, the dependency between most relevant and accurate classifiers is identified. The classifiers which were not chosen were combined toform the ensemble. This measure is supposed to ensurehigher performance and diversity of the ensemble.Various methods such as SA (Selection by Accuracy), SAD (Selection by accuracy and Diversity) and NS (No selection) algorithm were identified. [8] proposes various classification methods to classify the liver disease data set. The paper emphasizes the need for accuracy because it depends on the dataset and the learning algorithm. Classification algorithms such as J48, Naive Bayes were used to classify these diseases and compare the effectiveness, correction rate among them. The performance of the models was compared with accuracy and computational time. It was concluded that all the classifiers except naive bayes showed improved predictive performance. Multilayer perceptron shows the highest accuracy among the proposed algorithms. The paper [9] tries to solve the problem of food insecurity in Egypt. It proposes a framework which would predict the production, and import for that particular year. It uses Artificial Neural Networks along with multi-layer perceptron in WEKA to build the prediction. At the end of the process, we would be able to visualize the amount of production import, need and availability. Therefore, it would help to make decisions on whether food has to be further imported or not. The soil datasets in paper [10] are analyzed and a category is predicted. From the predicted soil category, the crop yield is identified as a Classification rule. Naïve Bayes and k-Nearest Neighbor algorithms are used for crop yield prediction. The future work stated is to create efficient models using various classification techniques such as support vector machine, principal component analysis.

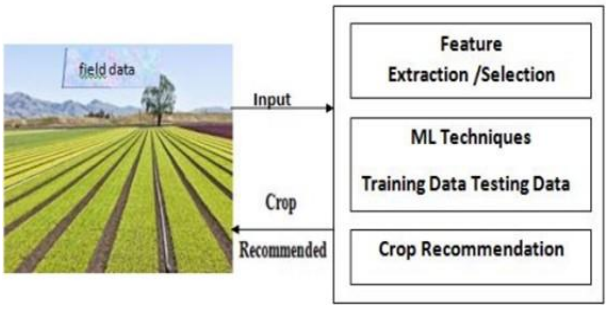
PROPOSED SYSTEM

We to eliminate the aforementioned drawbacks, we propose an Intelligent Crop Recommendation system, which takes into consideration all the appropriate parameters, including temperature, rainfall, location and soil condition, to predict crop suitability. This system is fundamentally concerned with performingthe primary function of Agriculture crop recommendation system, which is, providing crop recommendations to farmers algorithms. We also provide the profit analysis on crops grown in different states which gives the user an easy and reliableinsight to decide and plan the crops.

To overcome this method to implement machinelearning approach by user interface of GUI application

- Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would

be applied to extract patterns and to obtain results with maximum accuracy.



Architecture of Proposed system

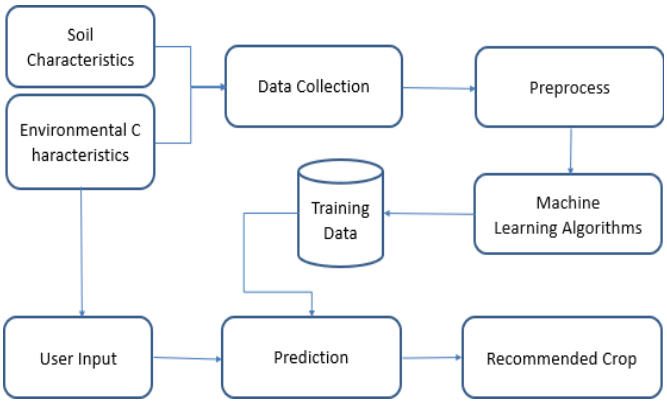
Advantages:

- The selection of crops & cropping system plays a major role in improving the productivity and profitability of the farmers. Crop

recommendation system thereby helps farmers during this decision-making process by considering various parameters such as temperature, rainfall, and seasons and agro-ecological situations.

- Crops are recommended based on soil, weather, humidity, rainfall, and other variables toincrease agricultural output. It benefits not just farmers, but also the country and helps to keep food costs down.

SYSTEM ARCHITECTURE



System Architecture

DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OFTHE PROPOSED MODEL/SYSTEM

Algorithm Explanation

Machine learning algorithms allow choosing the most profitable crop list. To predict the crop yield, selected Machine Learning algorithm such as K-Nearest Neighbor (KNN) is used.

Testing Methodologies

The program comprises of several algorithms which aretested individually for the accuracy. we check for the correctness of the program as a whole and how it performs.

Unit Testing

Unit tests focus on ensuring that the correct changes to the worldstate take place when a transaction is processed. The business logic in transaction processor functions should have unit tests, ideally with 100 percent code coverage. This will ensure that you do not have typos or logic errors in the business logic. The various modules can be individually run from a command line and tested for correctness. The tester can pass various values, to check the answer returned and verify it with the values given to him/her. The other work around is to write a script, and run all the tests using it and write the output to a log file and using that to verify the results. We tested each of the algorithms individually and made changes in preprocessing accordingly to increase the accuracy.

System Testing

System Testing is a level of software testing where a complete and integrated software is tested. The purpose of this test is to evaluate the systems compliance with the specified requirements. System Testing is the testing of a complete and fully integrated software product. and White Box Testing. System test falls under the black box testing category of software testing. Different Types of System Testing:

- Usability Testing - Usability Testing mainly focuses on the users ease to use the application, flexibility in handling controls and ability of the system to meet its objectives.
- Load Testing - Load Testing is necessary to know that a software solution will perform underreal-life loads.
- Regression Testing- Regression Testing involves testing done to make sure none of the changes made over the course of the development process have caused new bugs.
- Recovery Testing - Recovery testing is done to demonstrate a software solution is reliable, trustworthy and can successfully recoup from possible crashes.
- Migration Testing - Migration testing is done to ensure that the software can be moved from older system infrastructures to current system infrastructures without any issues.

Quality Assurance

Quality Assurance is popularly known as QA Testing, is defined as an activity to ensure that an organization is providing the best possible product or service to customers. QA focuses on improving the processes to deliver Quality Products to the customer. An organization has to ensure, that processes are efficient and effective as per the quality standards defined for software products.

Functional Test

Functional Testing is also known as functional completeness testing, Functional Testing involves tryingto think of any possible missing functions. As chat-bot evolves into new application areas, functional testing of essential chatbot components. Functional testing evaluates use-case scenarios and related business processes, such as the behavior of smart contracts.

Used Python Packages:

sklearn:

- In python, sklearn is a machine learning package which include a lot of ML algorithms.
- Here, we are using some of its modules like train test split, Decision Tree Classifier or Logistic Regression and accuracy score.

NumPy:

- It is a numeric python module which provides fast math functions for calculations.
- It is used to read data in NumPy arrays and for manipulation purpose.

Pandas:

- Used to read and write different files.

- Data manipulation can be done easily with data frames.

Python GUI:

- The Python GUI Project, or the PyGUI framework as it is more commonly known, is a simple API that enables developers to create user interfaces using native elementsfor Python applications.

K-NEAREST NEIGHBOR:

K-Nearest Neighbor can be used for both classification and regression. K-Nearest Neighbors is a non-complex algorithm which stores all the available cases and classifies new cases based on some similarity measure.The sample set is classified based upon the "closeness"that is the distance measure such as Euclidean distance or Manhattan distance.

Implementing K-Nearest Neighbor Algorithm

1. Importing the modules.
2. Creating Dataset.
3. Visualize the Dataset.
4. Splitting Data into Training and TestingDatasets.
5. KNN Classifier Implementation.
6. Predictions for the KNN Classifiers.
7. Predict Accuracy for both k values.
8. Visualize Predictions.

CONCLUSION

Farmers today need to pay close attention to all the things that can affect their crops. Throughout the growing period, people should seek advice and be observant of their own needs as well as those of the plants. Before planting crops, farmers should learn aboutand manage environmental factors that can affect crop growth. This helps them pick the best crop for theirsituation. Technology can help farmers analyze these factors and suggest the most suitable crops for cultivation.

The farmer provides feedback every month after thecrop is grown. This feedback helps improve the system'saccuracy over time as it self-trains with the data collected. With this system, a specialist is not required, and maintenance is minimal. Implementing this system won't cost the farmer any extra money.

In Sri Lanka, the average amount of land per person is smaller. According to sensor data, this system is ideal for both urban and rural settings and has an accuracy rate of above 95%. When more data is gathered and thesystem self-trains for around a year, accuracy will progressively increase. The system's accuracy and dependability were proved in tests that lasted 4-6 months. This method will serve as a model for other nations looking to introduce new agriculture consulting measures.

FUTURE WORK

- It is capable of many extra system functionalities. Today, it recommends a very suitable crop to be grown by taking relevant environmental parameters as inputs.
- But as the following level, the Automation component can be included as a feedback response mechanism. This can be changed to meet the needs of the farmer in order to adjust the humidity, water level, etc.
- Currently, all environmental factors are fed into the system as inputs, but as an added feature, an algorithm can be used to forecast one factor using data from two other factors. such that installing the sensors would cost less initiallyand be easy to maintain.

REFERENCES

1. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", International Conference on Smart Technologies and
2. Management for Computing, Communication, Controls, Energy and Materials, 2015.Haedong Lee and Aekyung Moon,
3. "Development of Yield Prediction System Based on Real-time Agricultural Meteorological Information", 16th International Conference on Advanced Communication Technology, 2014.
4. T.R. Lekhaa, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques", International Journal of Modern Trends in Engineeringand Science (IJMTES), 2016, Volume 03, Issue 10
5. Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar, "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction", International Journal of Computer Science Issues, Volume 9, Issue 3
6. Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia, SanyaBhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information, Embedded and Communication systems (ICIIECS).
7. AgroConsultant: Intelligent CropRecommendation System Using Machine Learning Algorithms. Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah.
8. Crop Recommendation System for Precision Agriculture S. Pudumalar*, E. Ramanujam*, R. HarineRajashree, C. Kavya, T. Kiruthika, J. Nisha.https://data.gov.in/g