IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124210

CROP PREDICTION USING SUPERVISED ML TECHNIQUE

Mrs.Saraswathy[1], Divya S[2], Jhansi S[3], Jija Bel M R[4]

[1]Assistant Professor, Department of Computer Science and Engineering, DMI College Of Engineering, Chennai, India

[2]B.E, Department of Computer Science and Engineering, DMI College Of Engineering, Chennai, India

[3]B.E, Department of Computer Science and Engineering, DMI College Of Engineering, Chennai, India

[4]B.E, Department of Computer Science and Engineering, DMI College Of Engineering, Chennai, India

Abstract- Agriculture provides most of the world's food and fabrics.Cotton ,wool and leather are agricultural product .Agriculture also provides wood for construction and paper products .These product, as well as the agricultural methods used, may vary from one part of the world to another . In general, agriculture is the backbone of India also plays an important role in the Indian economy. Now-a-days , food production and prediction are getting scarcity due to unnatural climate changes, which will adversely affect the economy of farmers by getting a poor yield and also help the agriculture analysis to remain less familiar with forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction .Supervised machine learning algorithm puts in the way to achieve it . The seed data of the crops are collected here, with the appropriate parameters like nitrogen, phosphorous, temperature, humidity, ph and rainfall content, which helps the crops to achieve the successful growth. To prevent this problem , agricultural sectors have to predict the crop from a given dataset using machine learning techniques. Also we are calculating the performance matrics.

INTRODUCTION

A.General

Crop yield prediction is one of the challenging tasks in agriculture. It plays an essential role in decision making at global, regional, and field levels. The prediction of crop yield is based on soil, meteorological, environmental, and crop parameters. Decision support models are broadly used to extract significant crop features for prediction. Precision agriculture focuses on monitoring (sensing technologies), management information systems, variable rate technologies, and responses to inter- and intravariability in cropping systems. The benefits of precision agriculture involve increasing crop yield and crop quality, while reducing the environmental impact.

B.Data Science

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. However, the definition was still in flux. The term "data science" was first coined in 2008 by D.J. Patil, and Jeff Hammerbacher, the pioneer leads of data and analytics efforts at LinkedIn and Facebook. In less than a decade, it has become one of the hottest and most trending professions in the market.Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.Data science can be defined as a blend of mathematics, business acumen, tools, algorithms and machine learning techniques, all of which help us in finding out the hidden insights or patterns from raw data which can be of major use in the formation of big business decisions.

C. Data Scientists:

Data scientists examine which questions need answering and where to find the related data. They have business acumen and analytical skills as well as the ability to mine, clean, and present data. Businesses use data scientists to source, manage, and analyze large amounts of unstructured data.

Required Skills for a Data Scientist**:**

- Programming: Python, SQL, Scala, Java, R, MATLAB.
- Machine Learning: Natural Language Processing, Classification, Clustering.
- Data Visualization: Tableau, SAS, D3.js, Python, Java, R libraries.
- Big data platforms: MongoDB, Oracle, Microsoft Azure, Cloudera.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124210

D.Artificial Intelligence

Artificial intelligence (ai) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. the term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by humans or animals. Leading AI textbooks define the field as the study of "intelligent agents" any system that perceives its environment and takes actions that maximize its chance of achieving its goals. Some popular accounts use the term "artificial intelligence" to describe machines that mimic "cognitive" functions that humans associate with the human_mind, such as "learning" and "problem solving", however this definition is rejected by major AI researchers.Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, and speech recognition and machine vision.AI applications include advanced web search engines, recommendation systems (used by Youtube, Amazon and Netflix), Understanding human speech (such as Siri or Alexa), self-driving cars (e.g. Tesla), and competing at the highest level in strategic game systems (such as chess and Go), As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology.Artificial intelligence was founded as an academic discipline in 1956, and in the years since has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success and renewed funding. AI research has tried and discarded many different approaches during its lifetime, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge and imitating animal behavior. In the first decades of the 21st century, highly mathematical statistical machine learning has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia.

LITERATURE SURVEY

Title : Crop yield forecasting on the Canadian Prairies using MODIS NDVI data
Author: M.S. Mkhabela, P. Bullock
Year : 2011
Normalised Difference Vegetation Index (NDVI) data derived from the advanced very high resolution radiometer (AVHRR) sensor have been extensively used to assess crop condition and yield on the Canadian Prairies and elsewhere, NDVI data derived from the new moderate resolution imaging spectro radiometer (MODIS) sensor have so far not been used for crop yield prediction on the Canadian Prairies. Therefore, the objective of this study was to evaluate the possibility of using MODIS-NDVI to forecast crop yield on the Canadian Prairies and also to identify the best time for making a reliable crop yield forecast. Growing season (May–August) MODIS 10-day composite NDVI data for the years 2000–2006 were obtained from the Canada Centre for Remote Sensing (CCRS). Crop yield data (i.e., barley, canola, field peas and spring wheat) for each Census Agricultural Region (CAR) were obtained from Statistics Canada. Correlation and regression analyses were performed using 10-day composite NDVI and running average NDVI for 2, 3 and 4 dekads with the highest correlation coefficients (r) as the independent variables and crop grain yield as the dependent variable. To test the robustness and the ability of the generated regression models to forecast crops grain yield, one year at a time was removed and new regression models were developed, which were then used to predict the grain yield for the missing year. Results showed that MODIS-NDVI data can be used effectively to predict crop yield on the Canadian Prairies.

Title : Remote Sensing of Environment
Author: . Becker-Reshef, E. Vermote
Year : 2010
Wheat is one of the key cereal crops grown worldwide, providing the primary caloric and nutritional source for millions of people around the world. In order to ensure food security and sound, actionable mitigation strategies and policies for management of food shortages, timely and accurate estimates of global crop production are essential. This study combines a new BRDF-corrected, daily surface reflectance dataset developed from NASA's Moderate resolution Imaging Spectro-radiometer (MODIS) with detailed official crop statistics to develop an empirical, generalized approach to forecast wheat yields. The first step of this study was to develop and evaluate a regression-based model for forecasting winter wheat production in Kansas. This regression-based model was then directly applied to forecast winter wheat production in Ukraine.

Title: Agricultural and Forest Meteorology
Author: Douglas K. Bolton. , Mark A. Friedl
Year : 2015
We used data from NASA's Moderate Resolution Imaging Spectro-radiometer (MODIS) in association with county-level data from the United States Department of Agriculture (USDA) to develop empirical models predicting maize and soybean yield in the Central United States. As part of our analysis we also tested the ability of MODIS to capture inter-annual variability in

DOI: 10.17148/IJARCCE.2023.124210

yields. Our results show that the MODIS two-band EnhancedVegetation Index (EVI2) provides a better basis for predicting maize yields relative to the widely used Normalized Difference Vegetation Index (NDVI). Inclusion of information related to crop phenology derived from MODIS significantly improved model performance within and across years. Surprisingly, using moderate spatial resolution data from the MODIS Land Cover Type product to identify agricultural areas did not degrade model results relative to using higher-spatial resolution crop-type maps developed by the USDA. Correlations between vegetation indices and yield were highest 65–75 days after greenup for maize and 80 days after greenup for soybeans. EVI2 was the best index for predicting maize yield in non-semi-arid counties (R2 = 0.67), but the Normalized Difference Water Index (NDWI) performed better in semi-arid counties (R2 = 0.69), probably because the NDWI is sensitive to irrigation in semi-arid areas with low-density agriculture.

Title : Plant Yield Prediction Model Using Firefly based Feature Selection with Modified Fuzzy Cognitive Maps
Author: D. Sabareeswaran, R. Gunasundari
Year : 2018
Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation. However, still the most agricultural fields are under developed due to the lack of deployment of ecosystem control technologies. Due to such issue 6, the crop production is not improved which affects the agriculture economy. Hence in this paper, a development of agricultural productivity is enhanced based on the plant yield prediction. Initially, different features such as plant images, soil characteristics, and weather factors are gathered and Firefly (FF) optimization algorithm is proposed for Feature Selection (FFFS). Then, the most selected optimal features are classified based on the Modified Fuzzy Cognitive Map (MFCM) algorithm for predicting the growth of plant yield. The predicted outcome is transmitted to the farmer's through smart phones which helps for identifying the growth of plant and improving the harvesting. The experimental results show that the effectiveness of the proposed technique can be compared with the other prediction techniques.

Title : Crop Yield Assessment from Remote Sensing
Author: Paul C. Doraiswamy, Sophie Moulin, Paul W. Cook
Year : 2022
Monitoring crop condition and production estimates at the state and county level is of great interest to the U.S. Department of Agriculture. The National Agricultural Statistical Service (NASS) of the U.S. Department of Agriculture conducts field interviews with sampled farm operators and obtains crop cuttings to make crop yield estimates at regional and state levels. NASS needs supplemental spatial data that provides timely information on crop condition and potential yields. In this research, the crop model EPIC (Erosion Productivity Impact Calculator) was adapted for simulations at regional scales. Satellite remotely sensed data provide a real-time assessment of the magnitude and variation of crop condition parameters, and this study investigates the use of these parameters as an input to a crop growth model. This investigation was conducted in the semi-arid region of North Dakota in the southeastern part of the state. The primary objective was to evaluate a method of integrating parameters retrieved from satellite imagery in a crop growth model to simulate spring wheat yields at the sub-county and county levels. The input parameters derived from remotely sensed data provided spatial integrity, as well as a real-time calibration of model simulated parameters during the season, to ensure that the modeled and observed conditions agree. A radiative transfer model, SAIL (Scattered by Arbitrary Inclined Leaves), provided the link between the satellite data and crop model. The model parameters were simulated in a geographic information system grid, which was the platform for aggregating yields at local and regional scales. A model calibration was performed to initialize the model parameters. This calibration was performed using Landsat data over three southeast counties in North Dakota. The model was then used to simulate crop yields for the state of North Dakota with inputs derived from NOAA AVHRR data. The calibration and the state level simulations are compared with spring wheat yields reported by NASS objective yield surveys.

PROPOSED SYSTEM

Active microwave remote sensing data at different frequencies can provide crucial information on crop morphology and conditions, thus effectively supporting agronomic management at different scales. Despite the ever-increasing availability of spaceborne platforms and the extensive research developed throughout more than two decades, some knowledge gaps still await to be filled toward operational use, dealing with SAR backscatter response to crop-specific features and seasonal dynamics, including the effects of agronomic practices. In this work, we used variance-based global sensitivity analysis (GSA) as a quantitative framework for investigating the sensitivity of X-band backscattering to agronomic and morphological features typical of two different crops maize and rice. To this end, we jointly exploited empirical data on crop status and growth, high-resolution Terra SAR-X (TSX) data, and microwave radiative transfer model (RTM) simulations. Phenology-informed simulations allowed us to quantify the contributions of different scattering mechanisms for the two crops under varying observation setups, to assess the sensitivity of X-band backscattering to morpho structural crop biophysical parameters (BPs) (and their interactions), and to evaluate the effects of crop biomass on backscatter across growth stages.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124210

A.Architecture Diagram:

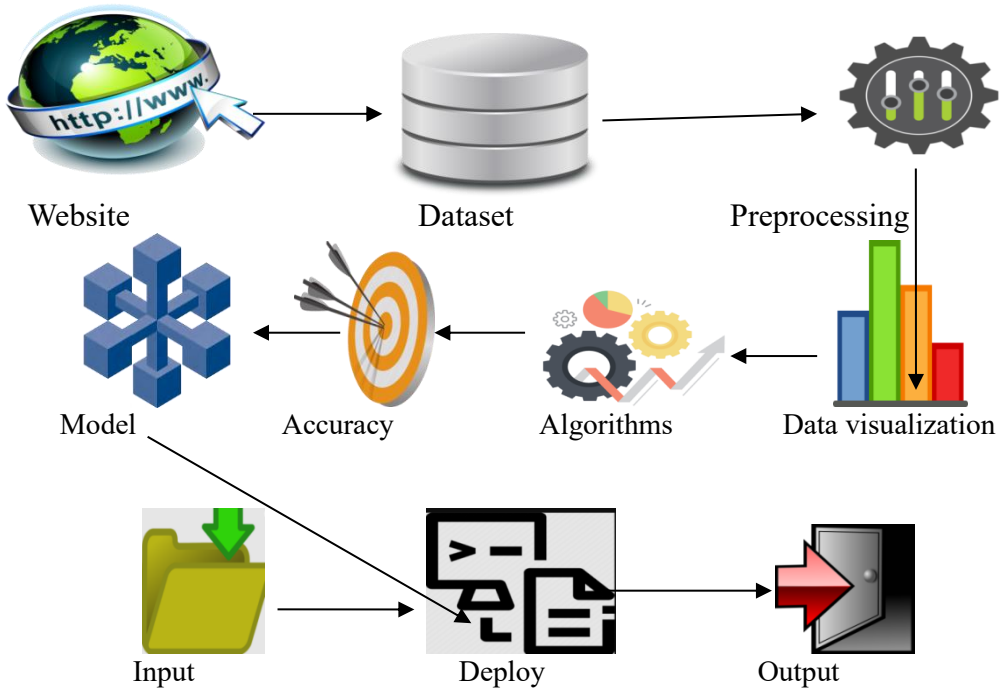


Fig.1. Architecture Diagram

2. Use Case Diagram:

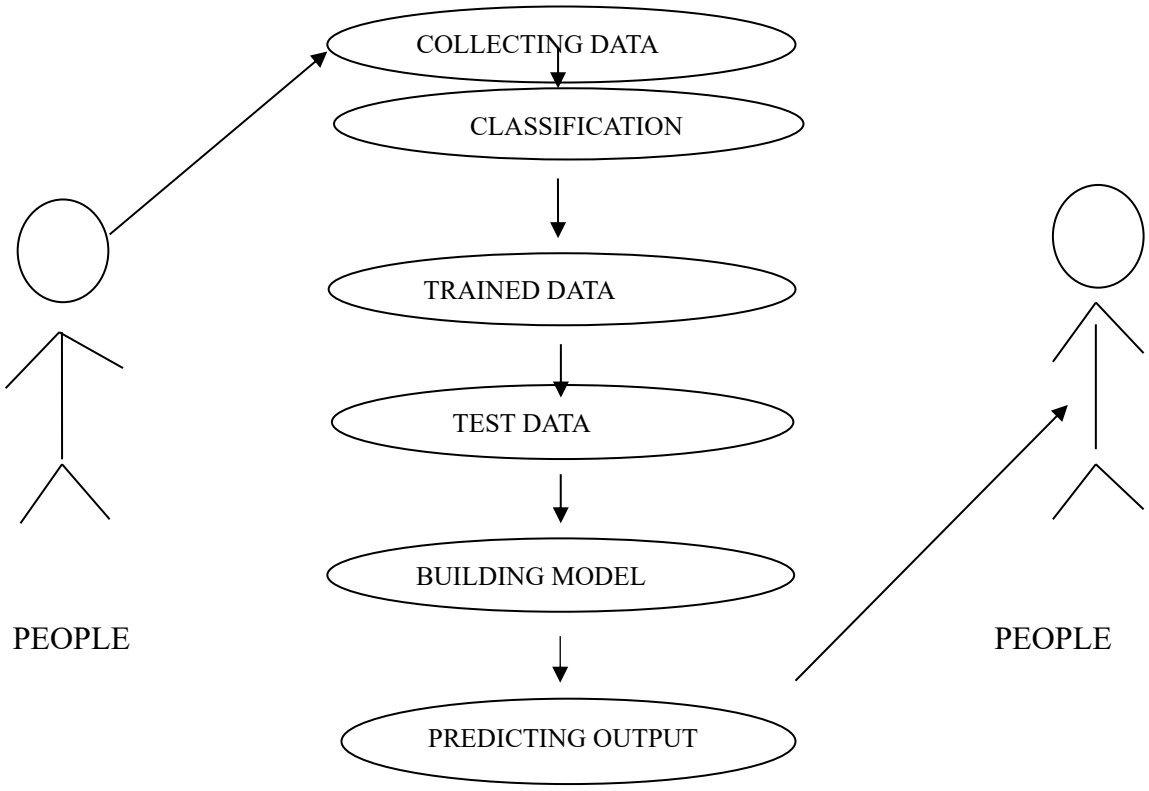


Fig.2. Use Case Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124210

MODULE DESCRIPTION

A.Data Preprocessing:

A number of different **data cleaning** tasks using Python's Pandas library and specifically, it focus on probably the biggest data cleaning task, **missing values** and it able to **more quickly clean data**. It wants to **spend less time cleaning data**, and more time exploring and modeling.Some of these sources are just simple random mistakes. Other times, there can be a deeper reason why data is missing. It's important to understand these different types of missing data from a statistics point of view. The type of missing data will influence how to deal with filling in the missing values and to detect missing values, and do some basic imputation and detailed statistical approach for dealing with missing data. Before, joint into code, it's important to understand the sources of missing data.

B.Data Virtualization:

This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

C. Algorithm Implementation:

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize.

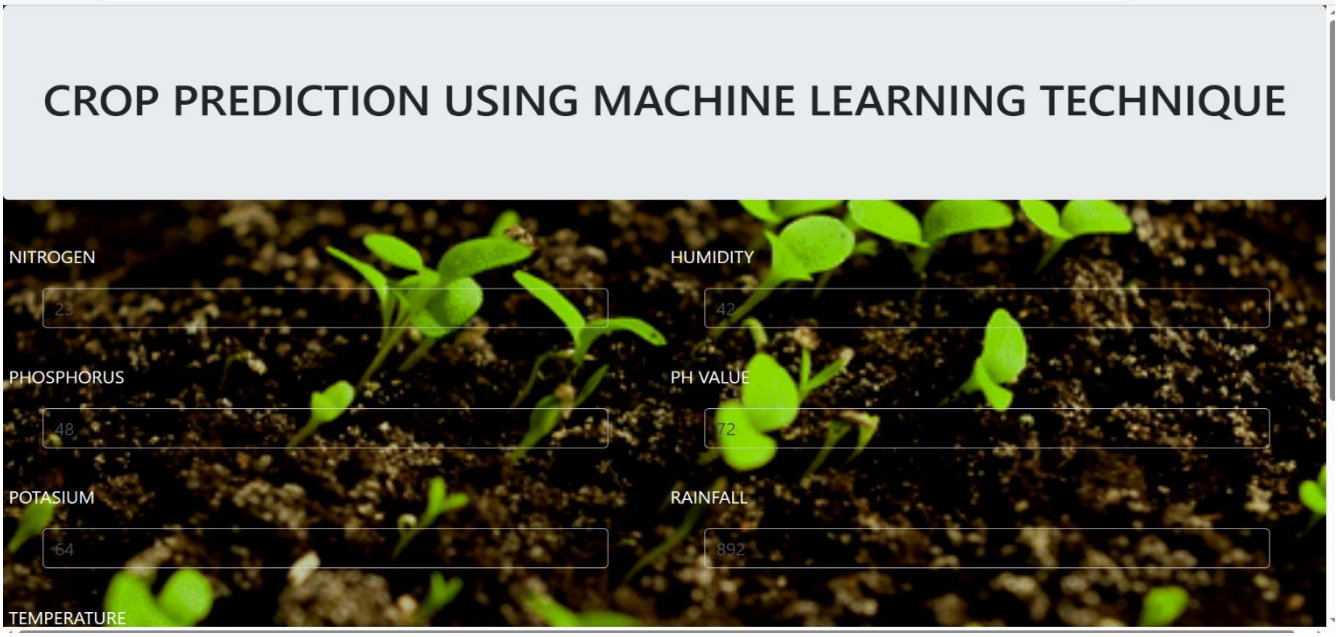
Naïve Bayes
Decision Tree Classifier
Ada Boost
Voting Classifier

D.Deployment.

After giving the input data among four algorithms decision tree classifier makes best choice for deploying the output during deployment the well nature of soil could be identified and crops can be grown accorndingly.

PROJECT OUTCOMES

Output

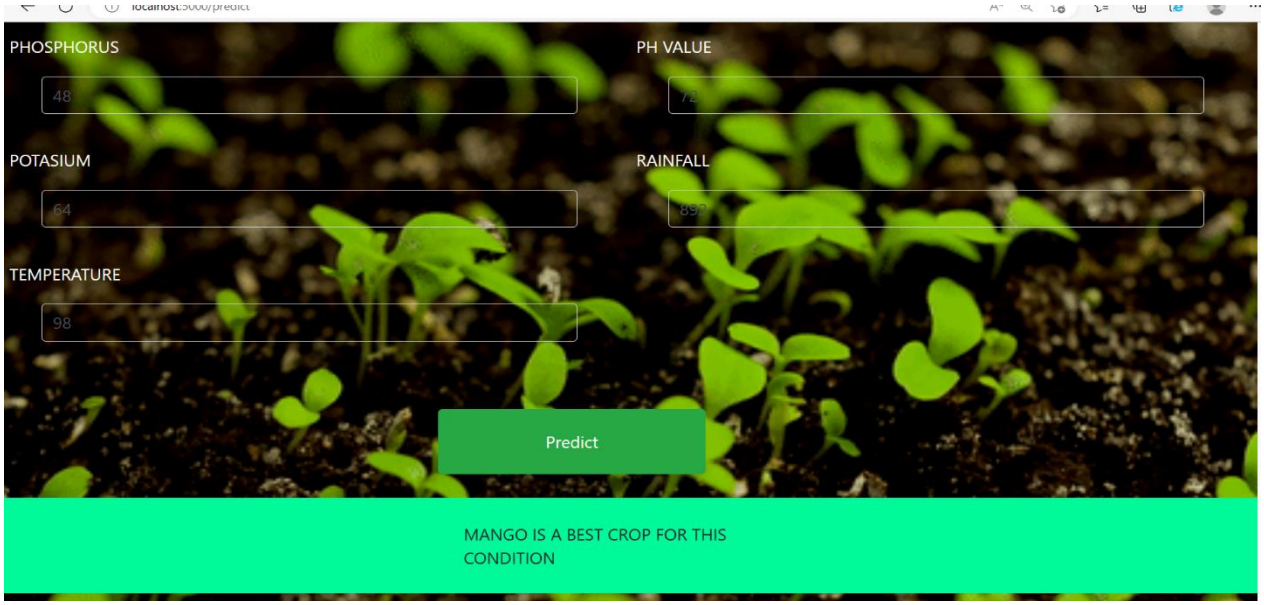


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124210



RESULT AND DISCUSSION

Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income. To reduce those type of loses we have developed a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the armers to take right decision in selecting the crop for cultivation such that agricultural sector will be developed by innovative idea.

CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set of higher accuracy score algorithm will be find out. The founded one is used in the application which can help to find the severity of cirrhosis in the patient.

REFERENCES

[1] Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning techniques IEEE 5th International Conference for Convergence in Technology (I2CT) pp 1-5 doi: 10.1109/I2CT45611.2019.9033611.

[2] Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning 2020 International Conference for Emerging Technology (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.

[3] Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop prediction using machine learning 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) pp 926-32 doi: 10.1109/ICSSIT48917.2020.9214190.

[4] Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S and Janani A P 2020 An effective crop prediction using random forest algorithm 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) pp 1-5 doi: 10.1109/ICSCAN49426.2020.9262311.

[5] Pande S M, Ramesh P K, Anmol A, Aishwaraya B R, Rohilla K and Shaurya K 2021 Crop recommender system using machine learning approach 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) pp 1066-71 doi: 10.1109/ICCMC51019.2021.9418351.

[6] Sellam V, and Poovammal E 2016 Prediction of crop yield using regression analysis Indian Journal of Science and Technology vol 9(38) pp 1-5.

[7] Bharath S, Yeshwanth S, Yashas B L and Vidyaranya R Javalagi 2020 Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction International Journal of Engineering Research & Technology (IJERT) NCETESFT – 2020 vol 8 Issue 14.

[8] Mahendra N, Vishwakarma D, Nischitha K, Ashwini and Manjuraju M. R 2020 Crop prediction using machine learning approaches, International Journal of Engineering Research & Technology (IJERT) vol 9 Issue 8 (August 2020).

[9] Gulati P and Jha S K 2020 Efficient crop yield prediction in India using machine learning techniques International Journal of Engineering Research & Technology (IJERT) ENCADEMS – 2020 vol 8 Issue 10.

[10] Gupta A, Nagda D, Nikhare P, Sandbhor A, 2021, Smart crop prediction using IoT and machine learning International Journal of Engineering Research & Technology (IJERT)