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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11473

Survey of Agriculture Production Optimization Engine Using Data Science with the Help of Machine Learning Predictive Model

Asha Mahiske[1], Dr. Tryambak Hiwarkar[2]

Student, Department of Computer Science and Engineering, WCEM, Nagpur, India[1]

HOD, Department of Computer Science and Engineering, WCEM, Nagpur, India[2]

Abstract: In the economic sector agriculture plays a vital role. Day by day the population is increasing on a large scale with this increases the demand of food. The early methods used by farmers are not sufficient enough to fulfill today's requirement, thus new methods are invented which in return brings employment for people. Machine learning Technology in agriculture has helped humans a lot such as identifying particular climate for particular crop similarly, it's soil type, pH value and water supply to the crop. The project consists of implementing a new method for different crop at similar time for larger productivity by predicting it accurately. In this paper we have studied various techniques presented by respected authors using Machine Learning and given below are the comparison between their respective technologies.

Keyword: Agriculture Production Optimization Engine, Palm oil, Crop Production, Crop Prediction, Machine Learning

INTRODUCTION

As we all know that agriculture depends largely on the nature of soil and the climatic conditions and many a times, we face unpredictable changes in climate like, non-seasonal rainfall or heat waves or fluctuations in humidity levels, etc. and all such events cause a great loss to our farmers and farming, because of which they are not able to utilize their agricultural land to it's fullest. So to solve all such problems, I have build a Machine Learning Model by the virtue of which we can help farmers, optimize the agricultural production, because this predictive model will help them understand that for a particular soil & given climatic condition, which crop will be best suitable for the harvest.
There are 7 key factors that I've taken into account which will help us in determining, exactly which crop should be grown and at what period of time, viz. Amount of Nitrogen, Phosphorus and Potassium in soil, Temperature in degree celcius, Humidity, pH and Rainfall in mm.

LITERATURE SURVEY

In this paper, the investigation has been done on crop yield prediction using Machine Learning approaches with special emphasis on Palm oil yield prediction. Under consideration of crop dataset different authors have used different Machine Learning Algorithms to predict some are given below.

A Comprehensive review of crop yield prediction using Machine Learning approaches with special emphasis on Palm oil yield prediction. In this project Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, Mohamad Anvar Kamaruddin and Nuzhat Khan have collectively worked on predicting palm oil yield. The Machine learning frameworks used offers a clear insight into the process by accessing vast sets of data and interpreting the obtained information. The model implementation of wide ranges of features like the difference in crop, location and intensity has also been observed. The selection of the features relies on the dataset's accessibility and the research objective. Altogether this paper focuses on indepth research on how crop yield prediction and palm oil yield prediction are linked to each other.

Development of Genetic Algorithm for Optimization of yield models in oil palm production. In this project Yousif Y. Hilal, Wan. I Shak, Azmi Yahya and Zulfa H. Asha'ari have collectively worked on a Mathematical Model for Palm Oil Yield using an Optimization Technique Genetic Algorithm(GA). This algorithm was used to produce fast and efficient solutions in correct time. The Genetic Algorithm with Correlation Analysis are easy to apply to a wide range to obtain Optimization Selection Models with high Precision Mathematical Models. The Sensitivity Analysis for the

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11473

variables choosen, one notices the effect of climate change, air pollution, and quality of land of oil palm areas simultaneously on several elements on land productivity.

Oil Palm Yield Forecasting based on Weather Variables using Artificial Neural Network. In this project Nadia Dwi Kartika, I Wayan Astika and Edi Santosa have collectively worked on Forecasting of Oil Palm Yield. This research used Secondary Data of Yield and Weather Variables available in Company Administration. It proposed feed forward Neural Network with Back Propogation Learning Algorithm to built a Monthly Yield Forecasting Model. The Optimization procedure of ANN Architecture obtained the best using 60 Neurons in input layer, five hidden layers and one Neuron in the output layer. The Training Data were from January 2005 to June 2008 while Testing Data were from July 2008 to December 2009.

Quality Prediction and Diagnosis of Refined palm Oil using Partial Correlation Analysis. In this project N A Rashid, K W Hoong, A Shamsuddin, N A M Rosely, M A M Noor, K W Jin, M H Lee and M K A Hamid have collectively worked on building a Model which gives a Quality Prediction and Diagnosis Tool using Partial Correlation Analysis. The Model aims on Predicting the Quality of Refined Oil and Diagnose the Crude Palm Oil and process variables. The Predictor Coefficient is developed using Partial Correlation Analysis while Control Chart is used to monitor the process behavior of both predicted and actual output. The monitored out of control behavior is then diagnosed using SPE-Contribution Plot to identify the faulty input variables, thus pre-treatment can be executed before the refining process.

Development of Geospatial Model for Predicting Metisa Plana's Prevalence in Malaysian Oil Palm Plantation. In this project S A Ruslan, Farrah Melissa Muharam, D Omas, Zed Zulkafli and M P Zambri have collectively worked on building a Model on Geospatial Technologies that can be used to obtain data in Rapid, Harmless and Cost Effective Manners. The Model utilized the technologies such as Land Surface Temperature(LST), Rainfall(RF), Relative Humidity(RH), and Normalized Difference Vegetation Index(NDVI), to examine climatic stresses that cause the outbreak, to predict the outbreak of Metisa Plana in Oil Palm plantation and to construct the relationship between Geospatial data and Metisa Plana Outbreak. Consistent day pattern was absent in the correlation between LST, RF, RH with Metisa Plana. Presence of Metisa Plana was not correlated with NDVI. In conclusion, this paper assures Metisa Plana's landscape ecology is possible with the utilization of Geospatial Technology and Temperature has been found to be the most important factor that influence the presence of Metisa Plana.

Examining the Forecasting Movement of Palm Oil Price using RBFNN-2SATRA Meta-heuristic Algorithm for Logic Mining. In this project Shehab Abdulhabib Alzaeemi and Saratha Sathasivam have collectively worked on RBFNN with different algorithms and the Logic Mining method for Forecasting constitute the most significant tools and techniques, which are used to demonstrate the Economic growth in the Country. Upon using monthly data spanning from January 2016 to March 2020 for the manufacturing of Palm Oil, the result mainly revealed that RBFNN-2SATRAAIS is the most accurate and efficient Model compared to RBFNN-2SATRAPSO and RBFNN-2SATRAGA in Forecasting the Price of Palm Oil. In the conclusion, the desire monthly price trend for Malaysia Palm Oil has been established with the highest accuracy.

Table 1: Features and Prediction Algorithms for Crop Yield Prediction

Reference	Objective	Feature	Prediction Algorithm	Performance
1	In depth research on how Palm Oil Prediction and Crop Yield Prediction are linked	Difference in crop, Irrigation Information, location, Soil Properties, Crop Land Information, Crop Management data, Fertilization Information, intensity, Climatic information, Land Surface Temperature Method(LSTM), Vegetation Indices, Historical Yield data, Ground water Characteristics	Linear Regression, Multiple Linear Regression, Multivariate Adaptive Regression Splines(MARS), K-Nearest Neighbor(K-NN), Support Vector Machine(SVM) and Support Vector Regression(SVR), Decision Tree(DT), Random Forest(RF), Extremly Randomized Tree(ERT), Artificial Neural Network(ANN), Convolutional Neural Network(CNN), Deep Neural Network(DNN)	

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11473

2	Prediction of Palm Oil Yield	Historical Yield Data, Crop land Information, Climatic Information, Air pollutants	Genetic Algorithm / Correlation Analysis(GA/CA)	R-squared: 0.948, MSE: 0.022
3	Palm Oil Yield Prediction	Climatic Information	Artificial Neural Network(ANN)	MAE: 0.5346, MSE: 0.4707
4	Quality Prediction and Diagnosis of Refined Palm Oil	Raw Crude Palm Oil data	Partial Correlation Analysis(PCA)	MSE less than 0.01
5	Prediction of Metisa Plana outbreak in Oil Palm Plantation	Land Surface Temperature Method(LSTM), Rainfall, Humidity and Normalized Difference Vegetation Index(NDVI)	Artificial Neural Network(ANN)	Accuracy: 95.42%
6	Palm Oil Price Forecasting	Monthly manufacturing data of Palm Oil	RBFNN-2SATRAAIS	Accuracy: 90.46%

Table 2: Summary of Advantages and Disadvantages of the Algorithms mostly used

Sr. no.	Algorithm Name	Key Features	Advantages	Disadvantages
1	Artificial Neural Network(ANN)	This network is inspired by the brain and essentially consists of input, output and hidden layers	It can perform complex nonlinear relationships between dependent and independent variables	It contains "black box" nature and greater computational burden
2	Convolutional Neural Network(CNN)	It consists of one input, hidden and an output layer whereas the hidden layers contain convolutional layers, ReLU layers, pooling layers, fully connected layers	A fully connected layer to convolution layer replacement indicates huge decrease in the number of parameters	A convolution is a considerably slower operation than Artificial Neural Network(ANN).
3	Deep Neural Network(DNN)	It utilizes several layers to gradually extract upper level properties from the raw input	It can address issues of over-fitting and local minima via an intensive optimization and activation process	Requires a high-end computer
4	Decision Tree	Decision Tree (DT) network consists of three nodes namely root node, decision node and leaf node	Decision Tree (DT) can be used in both classification and regression problem. It does not require normalization and scaling of data	This model has overfitting problem of the data which ultimately leads to wrong predictions

DOI: 10.17148/IJARCCE.2022.11473

5	Land Surface Temperature Method(LSTM)	It is the feed forward Neural Network (NN) with feedback from the input layer to the output layers of neurons	The sequence of data is modeled through the Recurrent Neural Network(RNN) such that each sample can be presumed to be dependent on previous ones	This network cannot process very long sequences when tanh or ReLU is used as an activation function
6	Multivariate Adaptive Regression Splines(MARS)	A method of non-parametric regression integrating a number of linear models to deal with nonlinearity and correlations between variables	It provides a versatile model capable of dealing with nonlinearities and linearities	It is restricted to handle large data and prone to over fitting issue
7	Random Forest(RF)	It consists of ensemble learning which utilizes the bootstrap and bagging process	It is less prone to over fitting problem than decision tree	It may change significantly by a small change in the data
8	Support Vector Machine(SVM)	The Support Vector Machine (SVM) constructs multi-dimensional boarders between data points in the feature space	It supports effective data grouping by optimizing the margin across categories employing kernel functions	Prone to over-fitting problems which depends on the kernel functions used for optimum classification

CONCLUSION

In this paper, we have encapasulated different techniques used by perspective authors in Machine Learning to predict the Palm Oil Yield Production with different datasets and with approaches. Most of the algorithms gives us an accuracy upto 90% in predicting it fast, correct and in correct time. There may be better techniques which may give us 100% accuracy and will be studied and added in future.

REFERENCES

1. A Comprehensive review of crop yield prediction using Machine Learning approaches with special emphasis on Palm oil yield prediction. By Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, Mohamad Anuar Kamaruddin, Nuzhat Khan on IEEE/ACCESS and received March 16 2021, accepted April 19 2021, date of current version May 3 2021.
2. Development of Genetic Algorithm for Optimization of Yield Models in Oil Palm Production. By Yousif Y Hilal, Wan Ishak, Azmi Yahya, and Zulfa H Asha'asri, Chilean journal of agriculture research, 78(2) : 228 – 237, 2018.
3. Oil Palm Yield Forecasting Based on Weather Variables Using Artificial Neural Network. By Nadia Dwi Kartika, I Wayan Astika, and Edi Santosa, Indonesian Journal of Electrical Engineering and Computer Science, 3(3) : 626 – 633, 2016.
4. Quality Prediction and Diagnosis of Refined Palm Oil Using Partial Correction Analysis. By NA Rashid, KW Hoong, A Shamsuddin, NAM Rosely, MAM Noor, KW Jin, MH Lee and MKA Hamid, In IOP Conference Series : Material Science and Engineering, volume 884, page 012018, IOP Publishing 2020.
5. Development of Geospatial Model for Predicting Metisa Plana's Prevalence in Malayasian Oil Palm Plantation. By S A Ruslan, F M Muharam, D Omar, Z D Zulkafi and M P Zambri, In IOP Conference Series : Earth asnd Environmental Science, volume 230, page 012110, IOP publishing 2019.
6. Examining the Forecasting Movement of Palm Oil Price Using rbfnn-2satra Metaheuristic Algorithms for Logic Mining. By Shehab Abdulhabib Saeed Alzaeemi, and Saratha Sathasivam, on IEEE/ACCESS, 9 : 22542 – 22557, 2021.