# CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering

by

GANDIKOTA JAYA NAGA DURGA PRASAD (38110151)



# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING SCHOOL OF COMPUTING

# SATHYABAMA

#### INSTITUTE OF SCIENCE AND TECHNOLOGY

# (DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH - 2022



#### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

#### BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Gandikota Jaya Naga Durga Prasad(38110151) who carried out the project entitled" CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM" under my supervision from October 2021 to May 2022

# InternalGuide

(Dr.C.Hemalatha)

# Head of the Department

Submitted for Viva voce Examination held on

**Internal Examiner** 

**External Examiner** 

# DECLARATION

I Gandikota Jaya Naga Durga Prasad hereby declare that the Project Report entitled CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM done by me under the guidance of Dr./Prof./Mr./Ms. Dr.C.Hemalatha (Internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering / Technology degree In COMPUTER SCIENCE AND TECHNOLOGY.

# **DATE:**

PLACE:

SIGNATURE OF THE CANDIDATE

# ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D**, **Dean**, School of Computing **Dr. L. Lakshmanan M.E., Ph.D.**, and **Dr.S.Vigneshwari M.E., Ph.D. Heads** of the Department of Computer Science and Engineering for providing menecessarysupport and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr./Mr./Ms **Dr.C.Hemalatha** for his valuable guidance, suggestions and constantencouragementpaved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Departmentof Computer Science and Engineering**who were helpful in many ways for thecompletion of the project.

### ABSTRACT

Fertilizer value updation has a positive practical significance for guiding agricultural production and for notifying the change in market rate of fertilizer to the farmer. 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. Different types of land condition. So the quality of the fertilizers are identified using ranking process. By this process the rate of the low quality and high quality fertilizer is also notified. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the crop for further.

# TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO
	ABSTRACT	1
1	INTRODUCTION	3
2	LITERATURE SURVEY	5
3	METHODOLOGY	14
4	CONCLUSION	23

FIGURE NO	FIGURE NAME	PAGE NO
1	GRAPH 1	19
2	GRAPH 2	19
3	SYSTEM ARCHITECTURE	19

#### CHAPTER 1

#### INTRODUCTION

In the world of developing technologies, the success of sharing information will help the agriculturists in realizing and developing their potential. The information sharing is that the valuable and timely information is being shared between agriculturists, either formally or informally. The willingness of information sharing refers to the open attitude among agriculturists. This open attitude determines the degree and scope of information sharing. Using web-technologies like html and css we build the web application, We create dataset by gathering data from multiple resources and place them in place which is used to predict the price of the fertilizers and reults are subjected to non-linear test later priorities are set and rankings are given to the list of fertilizers. Place information in our application and share that information to agriculturists whose data is collected and stored in the mysql server. we software to automatically send the updated information to the agriculturists in the form of text message so that agriculturists no need to go to near by towns and cities to know the updated information. We will be machine learning algorithms to predict the price of the fertilizers for the next two months. For prediction purpose we will be using machine learning algorithms to predict the crop for the further usage of the agriculturists. Further, a ranking process is applied for decision making in order to select the classifiers results.

#### **AIM & OBJECTIVE**

• Data set collection from various sources.

- Data parsing and cleansing technique is applied to make the raw data into processing data.
- The data collected is subject to machine learning system along with run time analysis makes an efficient fertilizer value updation system.
- Usage of Ensemble of classifiers makes the model more robust and efficient.
- Ranking technique used in the project helps us to make efficient decisions.
- Creating a web application for user registrations and collection of data.

#### CHAPTER 2

#### LITERATURE SURVEY

# [1]Title : A Review on Data Mining Techniques for Fertilizer Recommendation,2018

#### Authors : Jignasha M. Jethva, Nikhil Gondaliya, Vinita Shah

To keep up nutrition levels in the soil in case of deficiency, fertilizers are added to soil. The standard issue existing among the Indian agriculturists choose approximate amount of fertilizers and add them manually. Excess or deficient extension of fertilizers can harm the plants life and reduce the yield. This paper gives overview of various data mining frameworks used on cultivating soil dataset for fertilizer recommendation.

#### [2]Title : A Survey on Data Mining Techniques in Agriculture,2015

#### Authors : M.C.S.Geetha

Agriculture is the most critical application area especially in the developing nations like India .Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way. Data mining plays a important role in decision making on several aspects with agriculture field. It examines about role

of data mining in the farming field and their related work by a few authors in related to agriculture domain. It additionally talks about on various data mining applications in taking care of the several agriculture problems. This paper integrates the work of several authors in a single place so it is valuable for specialists to get data of current situation of data mining systems and applications in context to farming field.

#### [3]Title : AgroNutri Android Application,2016

# Authors : S. Srija, R. Geetha Chanda, S.Lavanya, Dr. M. Kalpana Ph.D

This paper communicates the idea regarding the making of AgroNutri an android application that helps in conveying the harvest particular fertilizer amount to be applied. The idea is to calculate the measure of NPK composts to be applied depend on the blanked proposal of the crop of interest. This application works depend on the product chosen by the farmer and that is taken as input, thus providing the farmers. The future scope of the AgroNutri is that GPRS can be included so that according to location nutrients are suggested. Further this application would be incorporated as a piece of the accuracy agriculture wherein sensors can be utilized to discover the measure of NPK present in the dirt and that sum can be deducted from the suggestion and giving us the exact measure of supplements to be added.

# [4]Title : Machine Learning: Applications in Indian Agriculture,2016 Authors : Karandeep Kaur

Agriculture is a field that has been lacking from adaption of technologies and their advancements. Indian agriculturists should be up to the mark with the universal procedures. Machine learning is a native concept that can be applied to every field on all inputs and outputs. It has effectively settled its ability over ordinary calculations of software engineering and measurements. Machine learning calculations have improved the exactness of artificial intelligence machines including sensor based frameworks utilized in accuracy farming. This paper has evaluated the different uses of machine learning in the farming area. It additionally gives a knowledge into the inconveniences looked by Indian farmers and how they can be resolved using these procedures.

# [5]Title : Impacts of population growth, economic development, and technical change on global food production and consumption,2011

Author: Uwe A. Schneider a, ît, Petr Havlik b, Erwin Schmid c, Hugo Valin b, Aline Mosnier b,c, Michael Obersteiner b, Hannes Bottcher b, Rastislav Skalsky´ d, Juraj Balkovicč d, Timm Sauer a, Steffen Fritz b

Throughout the following decades humanity will request more food from less land and water assets. This investigation evaluates the food production effects of four elective advancement situations from the Millennium Ecosystem Assessment and the Special Report on Emission Scenarios. partialy and jointly considered are land and water supply impacts from population development, and specialized change, and forests and agriculture demand request shifts from population development and economic improvement. The income impacts on nourishment request are registered with dynamic flexibilities. Worldwide farming area increments by up to 14% somewhere in the range of 2010 and 2030.Deforestation restrictions strongly impact the price of land and water resources but have little consequences for the global level of food production and food prices. While projected income changes have the highest partial impact on per capita food consumption levels, population growth leads to the highest increase in total food production. The impact of technical change is amplified or mitigated by adaptations of land management intensities

#### [6]Title : Brief history of agricultural systems modelling,2016

Author: James W. Jones a,\*, John M. Antle b, Bruno O. Basso c, Kenneth J. Boote a, Richard T. Conant d, Ian Foster e, H. Charles J. Godfray f, Mario Herrero g, Richard E. Howitt h, Sander Jansseni, Brian A. Keating g, Rafael Munoz-Carpena a, Cheryl H. Porter a, Cynthia Rosenzweig j, Tim R.Wheeler k

Rural frameworks science creates information that enables analysts to consider complex issues or take educated farming choices. The rich history of this science represents the decent variety of frameworks and scales over which they work and have been contemplated. Demonstrating, a basic apparatus in agrarian frameworks science, has been expert by researchers from an extensive variety of controls, who have contributed ideas and instruments over six decades. As agrarian researchers currently consider the "people to come" models, information, and learning items expected to meet the inexorably mind boggling frameworks issues looked by society, it is vital to check out this history

and its exercises to guarantee that we stay away from re-innovation and endeavor to think about all elements of related difficulties. To this end, outline here the historical backdrop of rural frameworks we demonstrating and distinguish exercises discovered that can help control the structure and advancement of up and coming age of farming framework apparatuses and techniques. Various past occasions joined generally innovative advancement in different fields with have unequivocally added to the development of farming framework demonstrating, including improvement of process-based bio-physical models of yields and domesticated animals, factual models dependent on verifiable perceptions, and financial streamlining and reproduction models at family unit and local to worldwide scales. Attributes of rural frameworks models have changed broadly relying upon the frameworks included, their scales, and the extensive variety of purposes that spurred their advancement and use by specialists in various controls. Late patterns in more extensive joint effort crosswise over establishments, crosswise over orders, and between people in general and private segments recommend that the stage is set for the significant advances in rural frameworks science that are required for the up and coming age of models, databases, learning items and choice emotionally supportive networks. The exercises from history ought to be considered to help stay away from barricades and entanglements as the network builds up this up and coming age of horticultural frameworks models.

# [7]Title : A Smart Agricultural Model by Integrating lot, Mobile and Cloud-based Big Data Analytics,2017

Authors : S.Rajeswari, K.Suthendran, K.Rajkumar.

In the cultivating field, the system models play a significant role to the enhancement of the agro-normal and money related conditions. In the proportions of benefits of the field and farm examinations to give the information and to recognize fitting and fruitful organization practices. It can recognize the organization to arrive managers and transversely over reality as long as the required soil, the board, environment, and money related information. Decision Support Systems (DSSs) use to make the information for the vermin the board, develop the officials. These systems are not using the impelled strategies to process the data. Thusly, use the adroit system thoughts to take the decisions for the issue. It expect a crucial activity in the comprehension of agronomic results, and their use as decision sincerely steady systems for farmers is extending.

[8]Title : An Overview of Internet of Things and Data Analytics in Agriculture: Benefits and Challenges, 2018

Authors : Olakunle Elijah, Tharek Abdul Rahman, Igbafe Orikumhi, Chee Yen Leow, Nour Hindia.

A blueprint of lot and DA in agriculture has been shown in this paper. A couple of zones related to the association of lot in agribusiness have been discussed in detail. The investigation of composing exhibits that there are clusters of work advancing being produced of lot development that can be used to increase operational efficiency and gainfulness of plant and creatures. The benefits of lot and DA, and open troubles have been identified and inspected in this paper. Iot is depended upon to offer a couple of benefits to the agribusiness division. Regardless, there are up 'til now different issues to be steered to make it moderate for close to nothing and medium-scale farmers. The key issues are security and cost. It is typical that as contention increases in the cultivating part

# [9]Title : Circulation Mode Selection Based on Cost Analysis,2017

# Authors: Xiurong Sun\*, Jingshan Zhang, Chenglin Wang, Tao Zhang

If every farmer and each average production base will join their optimal conditions in making cooperatives, it will accomplish economies of scale. Furthermore, producers will have an all the more favourable position in the plans with downstream firms (shipper or retailer).Second, the main customers of wholesale market are not inhabitants nearby who buy small quantities products but lower distributors or retailers. More redesigned transportation mode respects intensive attempt of new agrarian things, which prompts bolster the movement of new chain joint logistics and strengthen resource utilize and made logistics advantage quality. Refresh everything considered agrarian things spread. By then, regard the examination of gigantic worth control of standard things and achieve the mind blowing control to stream process.

# [10]Title: Support Vector Machine-based Fuzzy Self-learning Control for Induction Machines,2010

### Authors : Zongkai Shao

Using support vector machine (SVM) is to realize the self learning of fuzzy inference system (FIS), based on a fast modified varying metric method (MDFP) and a support vector machine identifier (SVMI), a SVM-FIS self-learning controller for the threephase induction machine adjustable speed system has been designed. The proposed controller is not only of the advantages that FIS does not depend on the plant model,

strong robustness, and adaptive self-learning ability, but also learning ability and generalization performance of SVM. The designed processes of SVM-FIS, MDFP, and SVMI algorithms have been described in details. Simulation results show the feasibility, correctness and effectiveness of the proposed control strategy, such as the excellent static and dynamic performances, and strong anti-interference ability.

# [11]Title : Machine Learning Facilitated Rice Prediction in Bangladesh,2015

# Authors: Mohammad Motiur Rahman, Naheena Haq, Rashedur M Rahman

In this examination, self organising map (SOM) was utilized to group the information relationship between the information factors. After that chisquare test strategy was utilized to set up the level of reliance between the related variable qualities. It was discovered that the day by day outrageous climate conditions, for example, most extreme and least fluctuation in temperature, precipitation, dampness and wind speed were the principle drivers of product development, yield and wine quality

# [12]Title : Support Vector Machine-Based Classification Scheme of Maize Crop,2017

#### Authors : Suhas S Athani, CH Tejeshwar

This paper says about, advancement of a mechanized framework to distinguish and group weeds from the products would be of extraordinary help and we have proposed a set-up that decreases labour. We have considered pictures of maize edits as the informational index. Separating surface highlights of the picture and applying SVM

(support vector machine) to arrange whether the given picture is a weed or a yield

brought about a precision of 82%. In this way, picture preparing is a proficient method to group the given picture about whether it is a weed or a yield. A similar classifier can be connected to recognize number of harvests like groundnut, paddy from weeds. The proposed framework gives a chance to investigate more about element extraction methods. Further research and usage may incorporate building up a mechanized equipment framework which could help in removal of weeds in the fields.

# [13] WITH MACHINE LEARNING ALGORITHMS FOR ESTIMATING WINTER WHEAT AREAS,2017

#### Authors : Y.Z. Pan2

we utilize different kernel functions in the CPPI models to depict the connection between fractional winter wheat area and MODIS EVI time series data. We tried three straight and non-direct kernel functions, including linear regression, artificial neural system, and support vector machine. The differences when utilizing various kernel functions are minor for areas with basic planting structure. For areas like DT where multiple crop types have comparative phenology cycles, ANN-CPPI is found to play out the best. The two crop types, to be specific winter wheat and rapeseed, can be separated well. These tests give elective answers for the uses of CPPI in mixed areas.

# CHAPTER 3 METHODOLOGY

#### **EXISTING SYSTEM**

The computational and data demands of structural price forecasting generally far exceed than what is routinely available in developing countries. Consequently, researchers often rely on parsimonious representations of price processes for their forecasting needs. Contemporary parsimonious form of price forecasting relies heavily on time series modelling. In time series modelling, past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. During the past few decades, much effort has been devoted to the development and improvement of time series forecasting models. Time series modelling requires less onerous data input for regular and up-to date price forecasting. Hence there is a need for better classification which would be an ensemble or hybrid classification model.

# DISADVANTAGES OF EXISTING SYSTEM

- Efficiency is low.
- More number of repeated work.

# **PROPOSED SYSTEM**

In proposed system, the data analysis technology is used to update the fertilizer rate change. 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. Different types of land condition. So the quality of the fertilizers are identified using ranking process. By this process the rate of the low quality and high quality fertilizer is also indimated. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the cost of the fertilizers for further. This project uses Ensemble of classifiers such as SVM, NAÏVE BAYES, KNN or hybrid classifier. In addition, this project uses Ranking technique.

### ALGORITHM USED

# 1. K-Nearest Neighbour (KNN) Algorithm:

K, N, as associate degree example is quite a foundation for teaching or teaching a lazy man concerning grace: it's a lot of concerning obtaining getting ready to wherever the calculation is postponed, and performance of the partition. K, N, machine learning algorithmic programs of the algorithm could be a terribly straightforward factor. and also the proximity arising from the category provide (class of K n) to the worth of the article (to proceed K N) is verified.

STEP 1: BEGIN

STEP 2: Input: D =

STEP 3: another instance of arranging x = (x1 ... Xn)

STEP 4: Count (xi, ci) d (xi, x) for each case composed.

Stage 5: Separate d (xi, x) from base to top, (I = 1 ... N)

STEP 6: x: Select K for instance close to Dkx

STEP 7: Scores x Dkx general classification

**STEP 8: Completion** 

#### 1. Naïve bayes Algorithm:

P(X) due to an earlier case. The technique relies on split Bayes associated with the conclusion of the first step on the assumption of free predictors. In the presence of the fixed function of defined limits I am in the presence of a simple categorizer Bayes too much foreign matter, and the other part of the bed. Even if it is the fruit of the well of the well - to shine and the properties of each other's special occasions, a companion of the opposites of one, or to confer the degree of his evil, whence it is said, `` which is good. Words for a Naive Bayes is a simple example, so that significantly terribly useful, and huge sets for the sake of knowledge. Simplicity is still attached to a more subtle kind of nice Bayes, the developer thought.



 $P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$ 

Above,

- P (c | x) offered the prophets the last mechanical chance (c, objective) (x, characteristic).
- P (c) is the main chance to watch out.

• P(x | c) is the capacity to anticipate the stage.

# 1. SVM Algorithm:

SVM upholds vector machines. For an informational index comprising of choices designed on an introduced name, the A SVM records models that anticipate another example order. Relegate other level/data displayed in classification 1. Assuming there are just two classifications, it tends to be shown as a paired SVM list. Here are a few kinds of SVM:

- SVM line
- Lines without SVM lines

# SVM Linear Classifier:

As far as enlistment, we will more often than not accept that the mentor gives a model at home. These information focuses are planned to overcome any issues. Hyperplane forecast is straightforwardly partitioned into two phases. The main thing to do when planning a hyperplane is to diminish the separation from the hyperplane to the closest information in two stages. The hyper-plane outline is displayed as the greatest hyper-plane.

# SVM Non-Linear Classifier:

Our data bundles are broadly appropriated all over the planet. Getting this data from totally various classes of hyperplants ought not be viewed as a decent choice. That is the reason Vapnik recommended making a nonlinear classifier utilizing a hyper-plane stunt. In the nonlinear SVM list, information focuses are relied upon to surpass the breaking point.

# Examples of SVM boundaries:

In this section, we will figure out how to pick the best hyperplan to execute. We will show you Category 2 data. The classes are displayed in triangles and circles.

# Case 1:

- Take a gander at the issue in Figure 2 and the data in the two unique classes. Presently we need to observe a decent hyper plane that can isolate the two classifications.
- For this situation, see Figure 1. on the option to see as the proper hyper plane In SVM, we attempt to build the distance between the hyper-plane and the closest information. This is known as an edge.
- Arrangement 1 is restricted, so it is more than the distance between the left and right sides of the example. So, our most elevated hyperplan edge will be "first".

GRAPH 1





Case 2:

• In Figure 2, we think about two distinct classes of media. Presently we need to find a decent hyperplane that can separate between the two classes.

Information for every class is circulated to the left or right. We will probably pick a hyperplan that can separate between classes for most extreme contrasts.

 For this situation, the choice limits are ordered, yet the limits of choice 1 demonstrate the most extreme distinction between \ bigtriangleup and \ bigcirc.

GRAPH 2



Fig.2

# ADVANTAGES OF PROPOSED SYSTEM

- Useful to people far away from towns/cities.
- Better time efficiency.
- Reduction of repeated work.

# SYSTEM ARCHITECTURE





#### SYSTEM REQUIREMENTS

#### HARDWARE REQUIREMENTS:

- System Pentium-IV
- Speed 2.4GHZ
- Hard disk 40GB
- Monitor 15VGA color
- RAM 512MB

#### SOFTWARE REQUIREMENTS:

- Operating System Windows XP
- Coding language Java
- IDE Eclipse

# MODULES

- User Login
- Metadata
- Data Pre-processing
- Prediction

### **User** login

This is the first activity, User needs to provide a correct contact number and a password, which user enters while registering, in order to login into the webpage. If information provided by the user matches with the data in the database table then user successfully login into the webpage else message of login failed is displayed and user need to reenter correct information. A link to the register activity is also provided for registration of new users.

### Metadata

All the main data used in the data set are initialized with the number to use in the algorithm it is like initializing all the details. In this metadata, we are going to initialize all the crop names with the numbers. This data makes us use the data easily in the algorithm. Hear the metadata of all the crops is given with a particular number. This number is not duplicated that is one number is given to one crop, the same number is not given to the other crop. This metadata consists of more than a hundred crops that grown all over India.

#### **Data Pre-processing**

Hear the raw data in the crop data is cleaned and the metadata is appending to it by removing the things which are converted to the integer. So, the data is easy to train. Hear all the data. In this preprocessing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data.

### Prediction

The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of agriculture field. This way we can help the farmers to grow the crop which gives them better yield.

# CHAPTER 4 CONCLUSION

This open methodology decides the level and degree of data trade. Enormous scope logical innovation can work on the exhibition of modern composts. The venture has fostered another strategy for estimating the cost of modern manures dependent on compost costs. The thought is to utilize a majority that isolates them to foresee. Utilizing a successive majority permits you to settle on better choices about speculations by utilizing various classifications. Also, the positioning system is utilized to settle on choices about the choice of results. The framework is utilized to uncover the cost of manure to further develop compost.

#### REFERENCES

[1] Manpreet Kaur, Heena Gulati, Harish Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Volume 99– No.12, August 2014.

[2] J. Meng, "Research on the cost of agricultural products circulation and its control under the new normal economic development," Commercial Times, no. 23, pp. 145147, 2016.

[3] A. Kaloxylos et al., "Farm management systems and the future Internet era," Comput. Electron. Agricult., vol. 89, pp. 130–144, Nov. 2012.

[4] N. N. Li, T. S. Li, Z. S. Yu, Y. Rui, Y. Y. Miao, and Y. S. Li, "Factors influencing farmers' adoption of new technology based on Logistic-ISM model-a case study of potato planting technology in Dingxi City, Gansu Province," Progress in Geography, vol. 33, no. 4, pp. 542-551, 2014.

[5] Y. Wang, "A neural network adaptive control based on rapid learning method and its application," Advances In Modeling and Analysis, Vol. 46(3), pp. 27-34,1994.