Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Crop Prediction System using Machine Learning

Prof. D.S. Zingade[1],Omkar Buchade[2],Nilesh Mehta[3],Shubham Ghodekar[4],Chandan Mehta[5]

Department of Computer Engineering,
All India Shri Shivaji Memorial Society's
Institute of Information Technology,
Kennedy Road- 411001.

Abstract —*India being an agricultural country, its economy predominantly depends on agriculture yield growth and allied agro industry products. In India, agriculture is largely influenced by rainwater which is highly unpredictable. Agriculture growth also depends on diverse soil parameters, namely Nitrogen, Phosphorus, Potassium, Crop rotation, Soil moisture, Surface temperature and also on weather aspects which include temperature, rainfall, etc. India now is rapidly progressing towards technical development. Thus, technology will prove to be beneficial to agriculture which will increase crop productivity resulting in better yields to the farmer. The proposed project provides a solution for Smart Agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent. Weather forecast data obtained from IMD (Indian Metrological Department) such as temperature and rainfall and soil parameters repository gives insight into which crops are suitable to be cultivated in a particular area. This work presents a system, in form of an android based application, which uses data analytics techniques in order to predict the most profitable crop in the current weather and soil conditions. The proposed system will integrate the data obtained from repository, weather department and by applying machine learning algorithm: Multiple Linear Regression, a prediction of most suitable crops according to current environmental conditions is made. This provides a farmer with variety of options of crops that can be cultivated. Thus, the project develops a system by integrating data from various sources, data analytics, prediction analysis which can improve crop yield productivity and increase the profit margins of farmer helping them over a longer run.*

Keywords-*Data Analytics, Prediction, Machine learning, Multiple linear regression, android application.*

I. INTRODUCTION

Agriculture is one of the most important occupation practiced in our country. It is the broadest economic sector and plays an important role in overall development of the country. About 60 % of the land in the country is used for agriculture in order to suffice the needs of 1.2 billion people. Thus, modernization of agriculture is very important and thus will lead the farmers of our country towards profit. [1]
Data analytic (DA) is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software.[2] Earlier yield prediction was performed by considering the farmer's experience on a particular field and crop. However, as the conditions change day by day very rapidly, farmers are forced to cultivate more and more crops. Being this as the current situation, many of them don't have enough knowledge about the new crops and are not completely aware of the benefits they get while farming them. Also, the farm productivity can be increased by understanding and forecasting crop performance in a variety of environmental conditions. Thus, the proposed system takes the location of the user as an input. From the location, the nutrients of the soil such as Nitrogen, Phosphorous, Potassium is obtained. The processing part also take into consideration two more datasets i.e. one obtained from weather department, forecasting the weather expected in current year and the other data being static data. This static data is the crop production and data related to demands of various crops obtained from various government websites. The proposed system applies machine learning and prediction algorithm like Multiple Linear Regression to identify the pattern among data and then process it as per input conditions. This in turn will propose the best feasible crops according to given environmental conditions.Thus, this system will only require the location of the user and it will suggest number of profitable crops providing a choice directly to the farmer about which crop to cultivate. As past year production is also taken into account, the prediction will be more accurate.

II. SURVEY DETAILS

JeetendraSheenoy, Prof. YogeshPingle[3]
This paper posits a solution to reduce the transportation cost and uses an IOT based approach in order to reduce the number of middle hops and agents between the farmers and end user which in turn will help the farmer. This paper turns out to be the motivation for our project. We incorporate the mechanisms implemented in the paper and additionally provide a prediction based mechanism in order to suggest crops which can maximize the profit.

@IJAERD-2017, All rights Reserved

Monali Paul, Santosh K. Vishwakarma, Ashok Verma[4]
In order to predict the yielding of the crops, the crops are analyzed and based on analysis they are categorized. This categorization is done based on data mining algorithms. This paper gives insight into various classification rules like Naive Bayes, K-Nearest Neighbor. Using this paper, we analyzed the classification rules and identified which will be appropriate for data set which we will be using in our project.

Abdullah Na, William Isaac,ShashankVarshney, EkramKhan[5]
This paper provides a smart phone based application which will measure the PH value of the soil, temperature and humidity in real time. The system uses a microcontroller block, sensing block and communication block. Sensors are employed in farm which can communicate with smartphones using Bluetooth in real time. This paper provides means of remote analysis of soil through various techniques. This paper encouraged us to look for various techniques through which we can transfer the data we will obtain from sensors for processing and eventually generating the output.

N. Hemageetha[6]
This paper discusses various data mining techniques like Market based Analysis, Association Rule Mining, Decision Trees, Classification and Clustering. It entirely covers Data Mining concept. Various data mining algorithms such as Naive Bayes classifier, J48, K-Mean are explained in this paper. It also provides classification of soil based on Naive Bayes, Genetic algorithm, Association Rule Mining. Eventually, it covers Clustering in soil database. This paper helped us in understanding and analysis of different data mining algorithms and classification mechanisms. This will prove to be extremely beneficiary while developing our project and will help in mining the dataset obtained from sensors employed remotely.

S.Nagini, Dr. T. V. RajiniKanth, B.V.Kiranmayee[7]
This paper theorizes an Explorative data analysis and discusses about designing of various predictive model. A sample data set is taken and various regression techniques are applied in order to identify and analyze the properties of each. Various regression techniques which are discussed in this paper are Linear, Multiple Linear, non-Linear, Logistic, Polynomial and Ridge regression. Using this paper, comparative study of various data analytics algorithm is obtained. This helps us to judge which algorithm best suits our proposed system.

AwanitKumar,Shiv Kumar[8]
This paper proposes a system for prediction of production of crops in the current year. In order to determine the crop production, it uses a data mining algorithm K-Means. This system also uses prediction mechanism in form of fuzzy logic. Fuzzy logic is a rule based prediction logic wherein a set of rules are applied on the land for farming, rainfall and production of crops. Using this paper, a clear insight of how K-Means can be used to analyze data sets is obtained. Similar to set of rules as they have applied in form of fuzzy logic, we will be applying the set of rules to predict which crop will yield maximum profit based on previous years cost of crops and current soil and weather data.

III. SYSTEM DESIGN

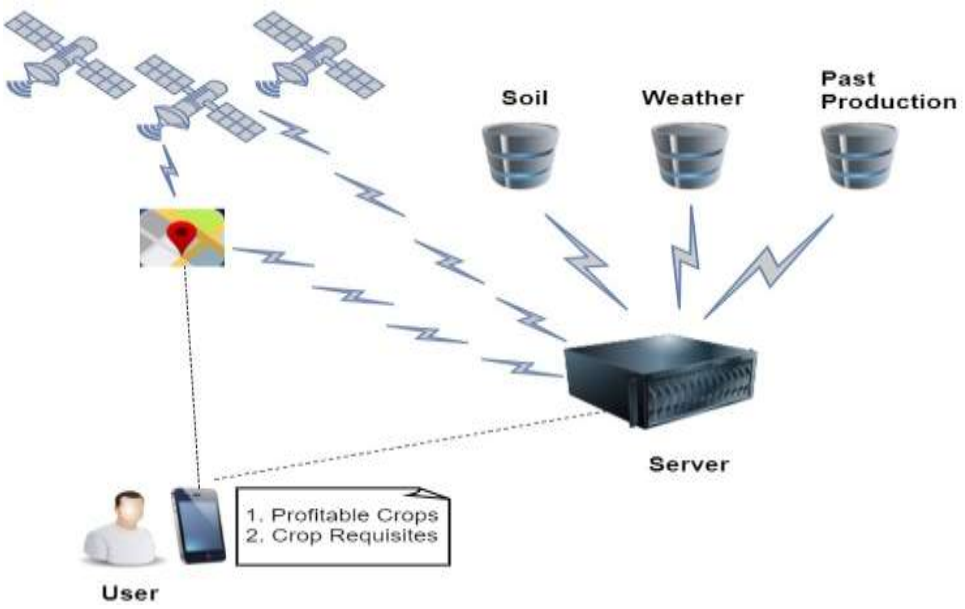


Figure 1. Proposed System Design

@IJAERD-2017, All rights Reserved

There is no existing system which recommends crops based on multiple factors such as Nitrogen, Phosphorus and Potassium nutrients in soil and weather components which include temperature and rainfall. The proposed system suggests an android based application, which can precisely predict the most profitable crop to the farmer. The user location is identified with the help of GPS. According to user location, the feasible crops in the respective location is identified from the soil and weather database. These soils are compared with past year production database to identify the most profitable crop in the current location. After this processing is done at server side, the result is sent to the user's android application. The previous production of the crops is also taken into account which in turn leads to precise crop proposition. Location is the only input for the extrapolation system. Depending on the numerous scenarios and additional filters according to the user requirement the most producible crop is suggested.

IV. ALGORITHMIC SURVEY

Regression: Regression analysis is a form of predictive modelling technique which investigates the association between a dependent (targets) and autonomous variable (s) (independent variables).

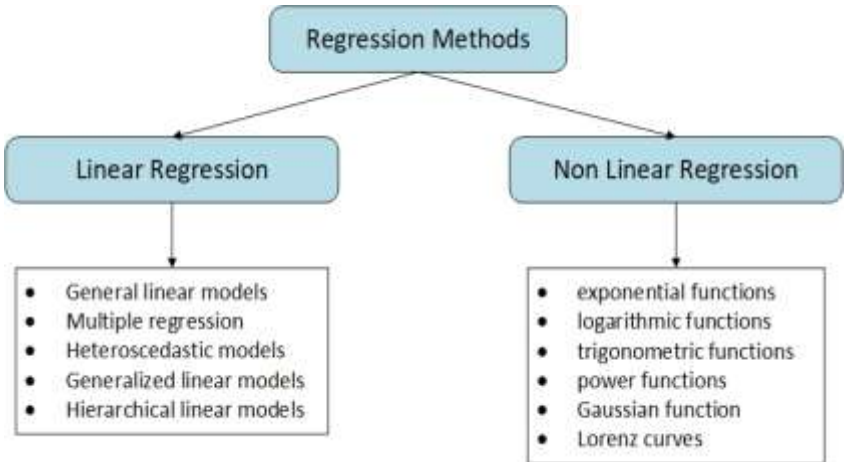
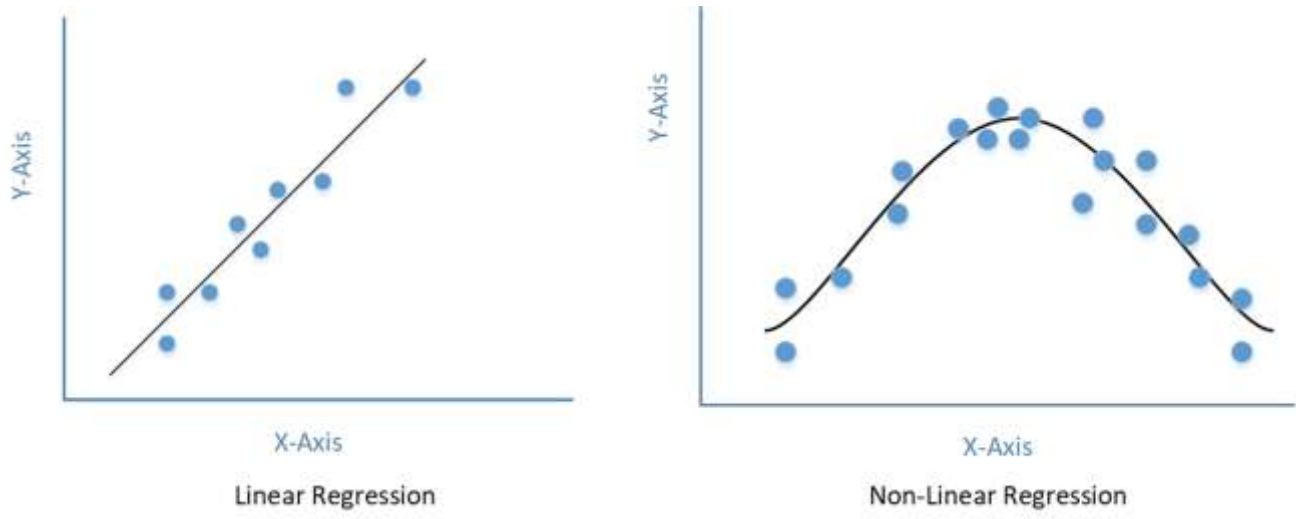


Figure 2. Algorithmic Classification

Linear Regression: Linear regression is a linear methodology for demonstrating the link between a scalar dependent variable y and one or more independent variables denoted X. The instance of solitary independent variable is called simple linear regression.[9]

Non Linear Regression: Nonlinear regression is a form of regression breakdown in which observational data are displayed by a function which is a nonlinear amalgamation of the model parameters and depends on one or more independent variables. The data is plotted by a technique of successive approximations.[10]



@IJAERD-2017, All rights Reserved

V. MULTI-LINEAR REGRESSINON

The difference between simple linear regression and multiple linear regression is that, multiple linear regression has (>1) independent variables, whereas simple linear regression has only 1 independent variable.
In this project, Multiple Linear Regression algorithm is used to predict the crops. Multiple Regression is an extension of simple Linear Regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables). For example, Multiple Regression to understand whether exam performance can be predicted based on revision time, test anxiety, lecture attendance and gender. Multiple Regression also allows you to determine the overall fit (variance) of the model and the relative contribution of each of the predictors to the total variance.

Formulae:
A Linear Regression model that contains more than one predictor variable is called a Multiple Linear Regression model. The following model is A Multiple Linear Regression model with two predictor variables, x_1 and x_2 .

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \in$$

Where,

$\beta_0, \beta_1, \beta_2$ … are coefficients of Multiple Linear Regression
x_{i1}, x_{i2} ... are independent variables.

The model is linear because it is linear in the parameters β_0, β_1 and β_2 . The model describes a plane in the three-dimensional space of $Y, x_1 \; and \; x_2$. The parameter β_0 is the intercept of this plane. Parameters β_1 and β_2 are referred to as partial regression coefficients. Parameter $\beta_1$represents the change in the mean response corresponding to a unit change in x_1 when x_2 is held constant. Parameter β_2 represents the change in the mean response corresponding to a unit change in x_2 when x_1 is held constant.

VI. PROPOSED SYSTEM ARCHITECTURE

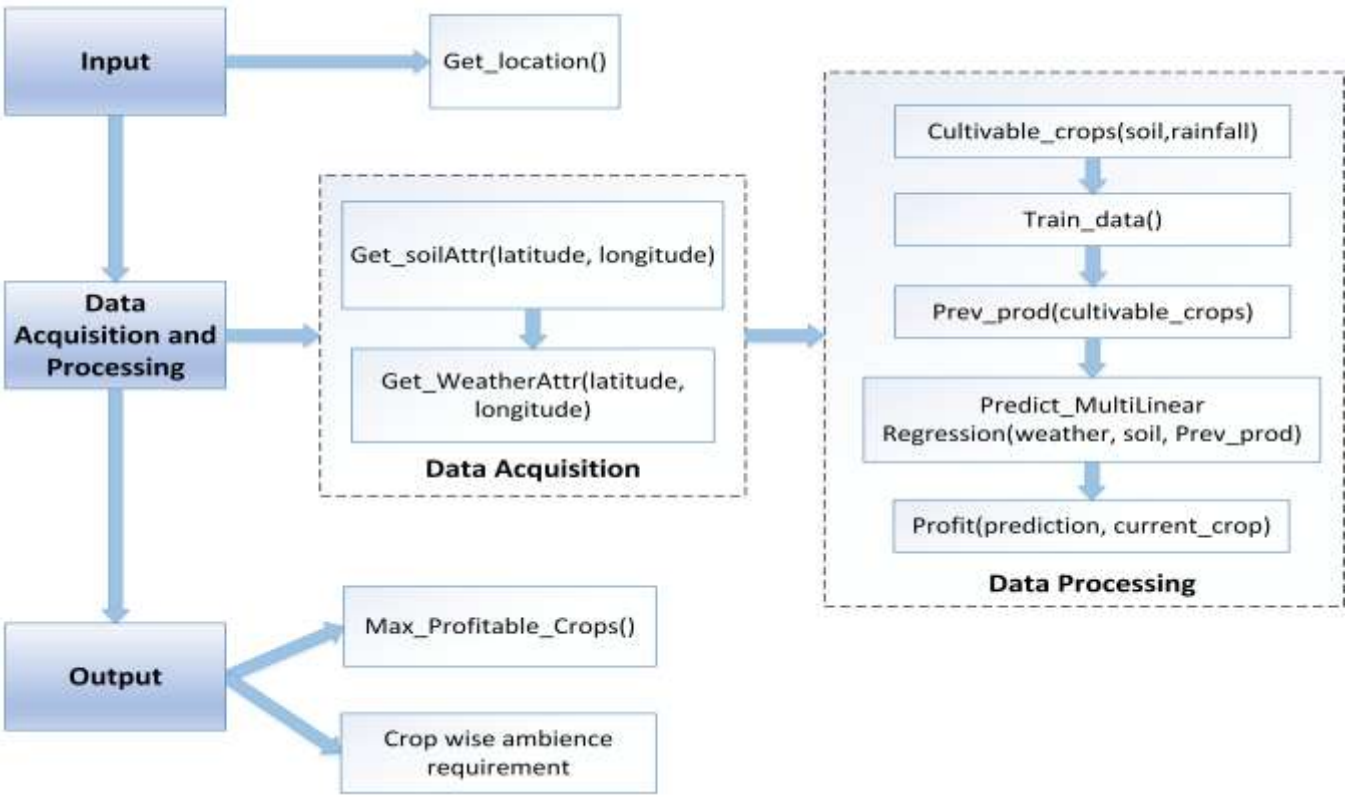


Figure 3. System Architecture

@IJAERD-2017, All rights Reserved

Input: The prediction of crop is dependent on numerous factors such as Soil Nutrients, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system.

Data Acquisition: Depending on the current user location, the system mines the soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set.

Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop. The prediction is based on past production data of crops i.e.: identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner.

Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

Set Theory:

S = {I, F_m, O, S, F}
I = {I_1} ……. *set of Input.*
I_1 = Location of user
F_m = { GetLocation(), GetAttributes(latitude, longitude),
GetSoil(), GetWeather(),
FeasibleCrop(soil,weather), PastProduction(),
ProfitableCrop(FeasibleCrops, PastProduction)
MaxProfitableCrops() } ……*Set of functions.*
Where, soil – N, P, K components
weather – Temperature and Rainfall values
O = {Crop predicted for given Location} ……. *Set of output.*
S= Correct prediction for High production and profit ……. *Success Condition*
F = Failure in prediction due to incorrect training data ……*Failure Condition*

VII. MATHEMATICAL REPRESENTATION OF ALGORITHM FOR PROPOSED SYSTEM

Train data:

$Y_1 = \beta_0 + \beta_1 x_{i2} + \ldots\ldots + \beta_p x_{ip} + \in_i$ for i=1,2, …, n

Where, $\beta_0, \beta_1, \beta_2 \ldots$are coefficients of Multiple Linear Regression
$x_{i1}, x_{i2} \ldots\ldots\ldots x_{ip}$ are independent variables.
X {weather attributes, soil attributes}
Y{production}
$Y = X\beta + E$
Y = Xβ + E
Y- production matrix X- attributes matrix B- Partial coefficient matrix E- error control
$\hat{\beta}$ = $(X'X)^{-1} X'Y$ ……Least Square Estimate
X' - Transpose X^{-1} - Inverse of Matrix

Prediction: $\hat{Y}$ = X$\hat{\beta}$
Result: res= $Y - \hat{Y}$

VIII. CONCLUSION AND FUTURE WORK

The proposed system takes into consideration the data related to soil, weather and past year production and suggests which are the best profitable crops which can be cultivated in theapropos environmental condition. As the system lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.

In the future, all farming devices can be connected over the internet using IOT. The sensors can be employed in farm which will collect the information about the current farm conditions and devices can increase the moisture, acidity, etc. accordingly. The vehicles used in farm like tractor will be connected to internet in future which will, in real time pass data to farmer about crop harvesting and the disease crops may be suffering from thus helping the farmer in taking appropriate action. Further the best profitable crop can also be found in light of the monetary and inflation ratio.

@IJAERD-2017, All rights Reserved

REFERENCES

[1]https://en.wikipedia.org/wiki/Agriculture

[2]https://en.wikipedia.org/wiki/Data_analysis

[3]JeetendraShenoy, YogeshPingle, "IOT in agriculture", 2016 IEEE.

[4] M.R. Bendre, R.C. Thool, V.R.Thool, "Big Data in Precision agriculture",Sept,2015 NGCT.

[5] Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining approach", 2015 International Conference on Computational Intelligence and Communication Networks.

[6]Abdullah Na, William Isaac, ShashankVarshney, Ekram Khan, "An IoT Based System for Remote Monitoring of Soil Characteristics", 2016 International Conference of Information Technology.

[7] Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R.Subhashri, "IOT Based Smart Agriculture Monitoring System", Feb 2017 IJRITCC.

[8] N.Heemageetha, "A survey on Application of Data Mining Techniques to Analyze the soil for agricultural purpose", 2016IEEE.

[9] https://en.wikipedia.org/wiki/Linear_regression

[10] https://en.wikipedia.org/wiki/Nonlinear_regression

[11] DhivyaB ,Manjula , Siva Bharathi, Madhumathi, "A Survey on Crop Yield Prediction based on Agricultural Data", International Conferencence in Modern Science and Engineering,March 2017.

[12] GiritharanRavichandran, ,Koteeshwari R S "Agricultural Crop Predictor and Advisor using ANN for Smartphones", 2016 IEEE,

[13]R.Nagini, Dr. T.V. Rajnikanth, B.V. Kiranmayee, "Agriculture Yield Prediction Using Predictive Analytic Techniques , 2nd InternationalConference on Contemporary Computing and Informatics (ic3i),2016

[14] Awanit Kumar, Shiv Kumar, "Prediction of production of crops using K-Means and Fuzzy Logic", IJCSMC, 2015

@IJAERD-2017, All rights Reserved