

# School of Computer Engineering & Technology



A Mini Project Report

*on*

“Crop prediction”

*by*

Kunal Sakhare (Exam Seat No.-T194054)

Arpit Gupta (Exam Seat No.- T194065)

Devashish Nannaware (Exam Seat No.-T194021)

Shubham Sable (Exam Seat No.-T194053)

*of*

**T.Y. BTech**

**Academic Year: 2019-2020**

# Contents

Sr. No.	Topic		Page No.
1	Abstract		3
2	Aim & Objective		4
3	Introduction		
	3.1	Introduction to Project	5
	3.2	Problem Statement	5
4	Proposed Model		
	4.1	Methodology	6
	4.2	Model Training	6
5	Dataset		
	5.1	Details of dataset	10
6	Results & Analysis (Output)		12
7	Conclusion		13
8	References		13

## **ABSTRACT**

This project focuses on using machine learning algorithm to predict the crop and related information. The required dataset is made and relevant information used for making it is taken from site '<https://data.world> › datasets › crops '.

We have used two different machine learning algorithms- linear regression and logistic regression. All these mentioned models successfully predict the crop 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 taken into account, the prediction will be more accurate.

## **AIM AND OBJECTIVE**

1. To classify the crop based on soil and state.
2. To predict the profit, yield and cost of cultivation for next years.

# INTRODUCTION

## 1.1 Introduction to Project

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.

Data analytics 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. 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.

## 1.2 Problem Statement:

Given the previous ten data for various crops according to the state which produces it and soil in which it is produced, find the crop's profit, cost of cultivation and yield for next years.

## **SURVEY**

### **2.1 Background Study:**

S.Nagini, Dr. T. V. RajiniKanth, B.V.Kiranmayee : 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.

## **PROPOSED MODEL**

### **3.1 Methodology**

The process methodology consists of the following steps – Data acquisition, data analysis, data wrangling and data modeling.

### **3.2 Model Training**

The process of training the model involves providing the learning algorithm with training data to learn from. To train the model Linear Regression and Logistic Regression were used. These algorithms are described briefly as follows:

#### **1) Regression:**

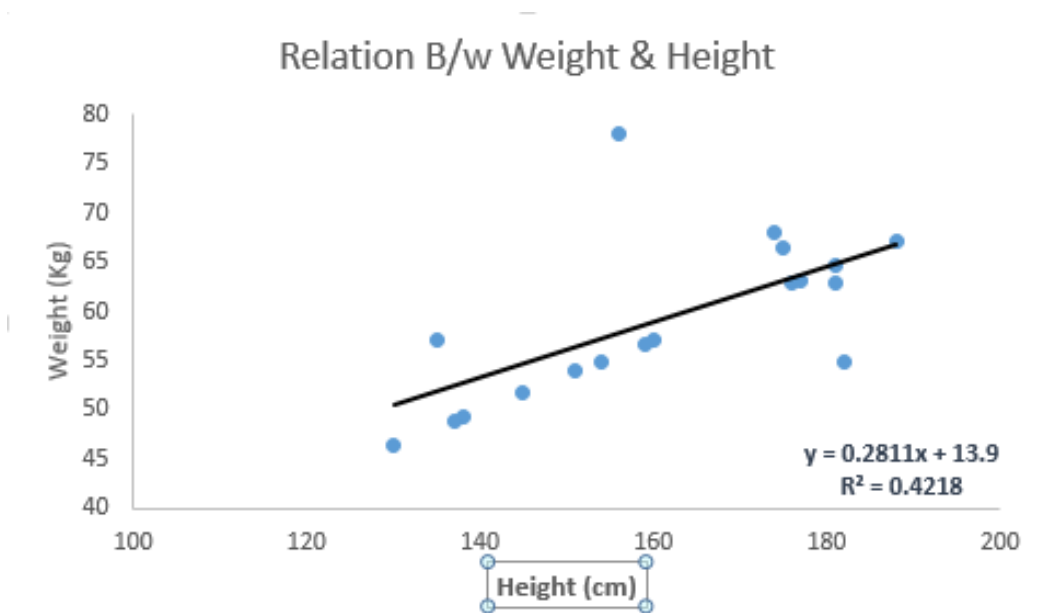
Regression analysis is a form of predictive modelling technique which investigates the association between a dependent (targets) and autonomous variable (independent variables).

#### **2) Linear Regression**

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Linear regression is used for predictive analysis. Linear regression is a linear approach for modeling the relationship between the criterion or the scalar response and the multiple predictors or explanatory variables. Linear regression focuses on the conditional

probability distribution of the response given the values of the predictors. For linear regression, there is a danger of overfitting.



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). The formula for linear regression is:  $Y = bX + A$ .



### 3) Logistic Regression

It is flow-chart like tree structure, where each internal node denotes a test on an attribute, each branch denotes an outcome of test, and each leaf node holds a class label. The topmost node in a tree is the root node. Given a tuple,  $X$ , for which the associated class label is unknown, the attribute values of the tuple are tested against decision tree. A path is traced from the root to a leaf node, which holds the class prediction for that tuple. The learning and classification steps of decision tree induction are simple and fast. Decision tree is useful because construction of decision tree classifiers does not require any domain knowledge. Their representation of acquired knowledge in tree form is easy to assimilate by users. Decision tree classifiers have good accuracy. The problem of constructing a decision tree can be expressed recursively. First, it is necessary to select an attribute to place at the root node, and make one branch for each possible value. The attribute with a highest information gain is selected as root node. Information gain is a measure of the effectiveness of the attribute in classifying the training data. When complexity of tree increases over fitting may occurs, which in turn reduces the accuracy of test data. To avoid over fitting pre-pruning and post-pruning methods are used. In pre-pruning, we stop early (while growing the tree) if gain is not statistically significant. While in post-pruning we grow full tree and remove the nodes based on cross validation.

## DATASET

### 4.1 Details of the dataset

The required dataset is made and relevant information used for making it is taken from site '<https://data.world › datasets › crops>'. We have created two datasets.

First dataset contains the soil and types of grown in that soil.

	SOIL	STATE	Rice	Cotton	Sugarcane	Wheat	Millets
0	Aluvial	Maharashtra	1	0	0	1	0
1	Black	Maharashtra	0	1	0	0	0
2	Red	Maharashtra	0	0	1	0	0
3	Laterite	Maharashtra	0	0	0	0	1
4	Mountain	Maharashtra	0	0	0	0	0

Fig.1 Overview of dataset

Second dataset is contains the data about cost of cultivation, yield and profit for various crops in various states from year 2008 to 2018.

	YEAR	STATE	CROP	YEILD	COST OF CULTIVATION	PROFIT
0	2008	Maharashtra	Rice	1.1	9743.05	5562.90
1	2009	Maharashtra	Rice	1.2	10523.56	7829.89
2	2010	Maharashtra	Rice	1.3	11568.99	6791.09
3	2011	Maharashtra	Rice	1.4	10908.52	8742.12
4	2012	Maharashtra	Rice	1.5	10783.33	8905.98
5	2013	Maharashtra	Rice	1.6	11031.59	3423.77
6	2014	Maharashtra	Rice	1.7	11873.21	6678.88
7	2015	Maharashtra	Rice	1.8	12523.24	7854.90
8	2016	Maharashtra	Rice	1.9	12323.17	12233.89
9	2017	Maharashtra	Rice	2.0	7891.07	2313.42
10	2018	Maharashtra	Rice	2.1	10674.34	5672.26

Fig.2 Overview of dataset

## RESULT AND ANALYSIS:

**Crop Recommendation System**

Maharashtra

Alluvial

Submit

Rice :  
Current Year Yield : [2.1]  
Current Year Profit : [3965.]  
Current Year Cost : [15523.]  
Cotton :  
Current Year Yield : [2.09]  
Current Year Profit : [12821.]  
Current Year Cost : [49057.]

The selected item is Maharashtra

Here we get two crop recommendations for growing crops in alluvial soil and Maharashtra state and showing their yield, profit and cost of cultivation. Based on that we can compare the yield, profit and cost of cultivation and decide the crop to be grown, in this case we can infer that if we want to grow crop which have less cost of cultivation we can choose rice or viceversa.

## CONCLUSION

We have successfully predicted the profit, yield and cost of cultivation for next year and recommended the crop based on soil and state using the machine learning algorithms. We have also successfully used Graphical user interface to take input and give output.

## REFERENCES

<https://www.geeksforgeeks.org › types-of-regression-techniques>

[www.ijaerd.com › papers › special\\_papers › RTDE007](http://www.ijaerd.com › papers › special_papers › RTDE007)

<https://towardsdatascience.com › 5-types-of-regression-and-their-properties...>