

CROP YIELD PREDICTION USING REGRESSION

D.Bhanu Kiran¹, J.Priyanka², K.Sri Poojitha³, Afridi Khan⁴

Abstract: As an agricultural country, India's economy is predominantly depend over production of yield from agriculture and agro- based products. By using Data Mining techniques we do research in the evaluation of crop yield. In agriculture, yield forecasting is a very important issue. The inspiration for this research comes from the agricultural point of view that the farmers have a lot of work to do manually. This work therefore deals with the automated approach to identification, which benefits farmers and decreases their efforts. With climate change in the world, crop mainstreams in the agricultural sector have been severely affected in previous decades performance. This paper uses regression models such as linear regression, decision tree regression, support vector regression, K- nearest neighbor regression models to discuss the prediction of crop production.

Keywords: Agriculture, Prediction, Regressions.

1. INTRODUCTION

India is one of the most densely populated nations with a wide range of fields and climatic conditions as well. Variable plants are grown in special parts of the country such as kharif, rabi, zaid crops due to particular climatic stipulations. But some farmers face serious problems in producing this vegetation due uncertain climatic conditions. Additionally, dry conditions add to these problems and can be serious threats as the soil plays a major role in the production of the crops. When a drought occurs the farmers hard work and effort goes to waste because manufacturing cannot be of no use. Work is underway to research these soils and determine the crops that are best grown in different areas with the aid of a variety of preconditions to support the farmers. Agriculture has ended up the backbone of our nation from historical instances. Weather prerequisites are very often different nowadays. Prediction of weather conditions, been stumbling block for growing vegetation. So we have to find a way or understand the background factors of the crop yield and the soil in order to understand growing crops as well fertilize plays one of the major factors to grow plants .Usage of fertilizer more or small quantities for the soil may also lose fertility and may not provide the desired crop yield. So, fertilizer becomes the main issue in it. The Weather conditions in the plays a prominent role in considerable imperative conditions for India as it can boost the Indian economy by assisting with the crop prediction because it plays a major role in the Indian economy. We have witnessed a number of advances in

technology in this era that are a response to many problems in terms of time, quality, cash or effort. Engineers are now collaborating with farmers to create a technological solution for agricultural elements. As the increasing increase in technological trends, however, there are problems as once the climatic preconditions change, all the conditions will be negative for the farmers, leading to economic loss as it plays an important role. Research shows that inspecting all the attributes such as soil, rainfall and so on, we can predict the type of crop that is suitable for development in specific locations. These facts help to improve these crops in the future and can support farmers. In some parts of us a demand for a particular crop may also be extra and the soil may be more likely to produce that particular crop, but due to some droughts and also lack of rain, it may also make the soil at appropriate for growing that particular crop. Analyzing and forecasting such plants is easier for such conditions because they may know what specific crop can be produced. Famer hard work and effort is not wasted either. Improving the quality of the plant is the main aim of gaining a clear picture understanding of the crop yeild through the data mining methods. Rentability and sustainability are the main goals of precision agriculture.

2. RELATED WORK

The Data Mining methodology ' overall goal is to extract the knowledge from a knowledge set and switch it into a plain framework for additional use. This paper analyzes the assembly of crop yields supported the info collected. So as to maximise plant production, the info mining methodology was accustomed predict crop yield. There are 2 distinct crop models:

1. Simulation model

2. Regression model

1. Simulation model portrays the scientific connections natural for the set from past investigations. This technique will generate outputs below numerous conditions rough in depth data accustomed establish and check the model.

Here in this crop yield prediction the farmer must offer the world of land, region, economic standing and town. One selecting the town parameter supported altitude, line of longitude and latitude climatological

information will be reflected from crop mental item. The element decision module is chargeable for set determination of quality from crop mental article for study learning. The harvest metal article is conveys with it ranch data like district- id, area name, soil-type, waterph, precipitation, moistness, daylight, land information, base conjointly incorporates the examples of yield with comparing ranch information, ecological parameter and pesticides data. The above data is for the ranch data and harvest data, for example, crop sort seed type.

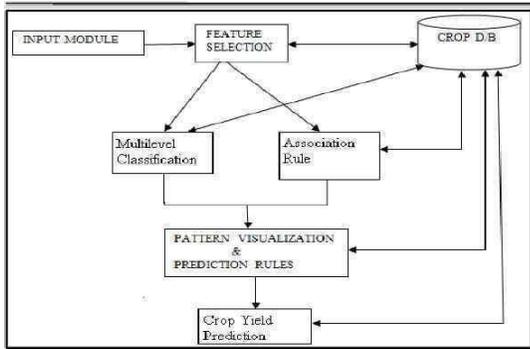


Fig1: Crop prediction using data mining techniques.

The input module will select the feature selection weather it will take multilevel classification or affiliation rule. Natural parameter, and pesticides information. When set decision of trait, the data goes to characterization and affiliation rule for gathering comparative substance. At that point expectation rules will be applied to yield of pack to instigate prompts terms of harvest, substance and cost..

3. METHODOLOGY

Usage is a procedure of arranging an article into a specific class dependent on the preparation informational collection that was utilized to prepare the classifier. We feed the classifier with informational index so we can prepare it to recognize related articles with as most ideal exactness as. We have utilized diverse relapse strategies like straight relapse, choice tree relapse, K-closest Regression, Support vector Regression.

Linear Regression:

Linear Regression is an algorithm based on supervised learning for machine learning. It is playing out an undertaking of relapse. Relapse models a prescient objective worth dependent on factors of indicators. It is generally used to decide the connection among factors and expectation. Distinctive relapse models contrast based on the sort of connection between anticipated factors and indicators, they are considered and the number of predictor variables used. A linear relationship

is the simplest mathematical relation between two variables u and v. In a relationship between cause and effect, the predictive variable is the source, and the effect is the expected variable.

Decision Tree Regression:

Decision tree constructs regression (or) tree structure classification model. This breaks down the dataset into smaller and smaller sub- sets while at the same time gradually creating a related decision tree. The algorithm for decision-tree falls within the category of supervised learning algorithms. It functions for both continuous and categorical output variables. The branches/edges are the node outcome and the nodes have either:

1. Conditions [Decision Nodes]
2. Result [End nodes]

Decision tree regression examines an object's characteristics and trains a model in a tree's structure to predict information for realistic continuous production in the future.

Continuous output means the output(or)result is not discrete, i.e. it is not only defined by a single set of numbers or values that is identified.

K-Nearest Neighbors:

K closest neighbors is a simple algorithm which stores all possible cases and predicts the solution based on a measure of similarity and dissimilarity (e.g., distance functions). Previously KNN was used as Non parametric tool for numerical estimation and pattern recognition. A simple implementation of KNN regression is to calculate the average K nearest neighbors numerical target. Another approach uses the K nearest neighbors' weighted average of the inverse distance. Regression of KNN uses the same distance function as classification of KNN.

S.NO	ALGORITHM	PRODUCTION
1.	Linear regression	0.087
2.	Decision tree	-1.200
3.	K-Nearest	0.35

Table1: Comparison of production rate for different algorithms.

IMPLEMENTATION

Step1: Crop dataset Collection

Step2: Pre-processing of dataset

Step3: Visualizing of data

Step4: Apply regression algorithms

Step5: Choose best regression algorithms

Step6: Calculate the production

Data Set Description:

We gathered the dataset from the Government administration site. It contains parameters like State Name, District Name, Crop Year, Season, Crop, Area, Production.

State Name: Data of states collected from states in India.

District Name: The data was collected from each of the districts in each of the state in India. **Crop Year:** Data was taken from the year 1997- 2015.

Season: Data was collected for each season like kharif, Rabi etc.

Crop: Plants such as paddy, wheat, cholam, ragi etc from a particular area.

Area: Total area of agricultural plant area in acres.

Production: Crop production in metric tons for the specified year.

Result:

We have visualized the area of production of crops for all over India

Crop_Year	Production	Area
1997	6.584859e+07	38786467.00
1998	7.876122e+07	38716952.00
1999	7.791707e+07	37858642.00
2000	8.209270e+07	43198618.00
2001	9.486370e+07	42745829.00
2002	7.146340e+07	46795168.29
2003	8.783597e+07	43181631.06
2004	8.417252e+07	42664655.23
2005	8.508719e+07	43519263.58
2006	9.167823e+07	43263797.59
2007	9.206461e+07	42826154.21
2008	9.526127e+07	43397987.00
2009	8.735634e+07	40583697.00
2010	9.517621e+07	41766988.50
2011	1.420929e+08	42811883.74
2012	1.083395e+08	47765615.00
2013	1.820729e+08	46485966.00
2014	1.848182e+08	39888873.00
2015	5.893687e+06	3922669.00

Bar plot of the above data
Line plot of the above data

Fig1: Production of crops and the area.

We have taken x as the state name and y as the crop name and z as the district name and we have calculated the production of the area where crops are growing very fast.

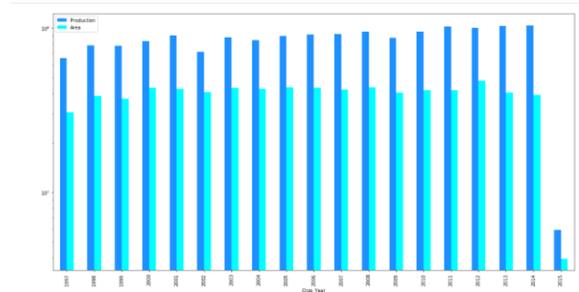


Fig2: Visualization of area and production of a crop for a state.

4. CONCLUSIONS

Systematic efforts are being made in this study to design a system that results in crop production prediction. Machine learning Regression algorithms were learned and tested on different measures during this research. Experiments and simulations are performed on the Database of crop yield. Experimental results assess the system's acceptability.

The proposed framework considers the information identifying with the yield of season, area and past year and recommends which are the best beneficial harvests that can be developed in the right natural condition.

Since the program records every potential yield being developed, it permits the rancher to figure out what plant to develop. This strategy additionally considers past information yield that will enable the rancher to pick up understanding into advertise request and the cost of various harvests. Since all out plant types under this program will be ensured, as well as farmers will find out about the harvest that may never have been delivered

5. FUTURE WORK

In the future, IOT can be used to connect all farming devices to the internet. In the farm, the sensors can be utilized to gather data about the present homestead conditions and gadgets can in this manner increment the humidity, acidity, etc. Later on, on-farm vehicles, for example, tractors will be associated with the web which, progressively, will pass information on crop collecting to ranchers and the sickness harvests may endure therefore, helping the farmers. In addition taking into account the fiscal and expansion proportion, the best productive yield can likewise be effectively found.

6. REFERENCES

1. <https://ieeexplore.ieee.org/document/8698087>
2. http://www.periyaruniversity.ac.in/ijcii/issue/Vol6No4Mar2017/M5_PID0370.pdf
3. <http://ijraset.com/files/serve.php?FID=22039>
4. https://www.academia.edu/9421293/A_survey_on_Data_Mining_Techniques_for_Crop_Yield_Prediction
5. http://www.ijaerd.com/papers/special_papers/RTDE007.pdf
6. <https://pdfs.semanticscholar.org/3376/e91c3a77a547ce51cfe4a2e68ea6f35ffe63.pdf>
7. Niketa Gandhi et al, "Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques", IEEE International Conference on Advances in Computer Applications (ICACA), 2016.
8. Dakshayini Patil et al, "Rice Crop Yield Prediction using Data Mining Techniques: An Overview", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May 2017.