Yield Prediction using Machine Learning for Agriculture Insurance

Rajneesh Kumar [1], Rajendra Prasad Mahapatra [2*]

[1, 2] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Delhi-NCR Campus, Ghaziabad, India

*Corresponding Author Email: mahapatra.rp@gmail.com

Abstract

India has achieved self dependency in food grain production by using modern technologies and better scientific approaches for cultivation. Beside this framers are less dependent on monsoon due to better irrigation facilities. Indian government provides various subsidies on seed, electricity and other agricultural equipment's to increase the income of the farmers. Indian government started various insurance schemes like PMFBY (PRADHAN MANTRI FASAL BIMA YOJANA) so that in spite of these efforts if due to some natural calamities like flood, drought or hail rain etc. farmers suffers losses then agriculture insurance is acting as a helping hand to manage the risk in yield loss to the farmers. We can say agriculture insurance is a risk management tool where the crop production risk is bear by government at a minimal cost. Millions of farmer's are getting benefited by this schemes still agriculture insurance has not reached to maximum number of the farmers due to lack of awareness. The paper proposed a scientific approach using machine learning to calculate the loss to the farmers. So that they can be benefited at the most. The accuracy of the model comes out be 91%.

Keywords

Agri Insurance, Crop Insurance, Crop Yield, Machine Learning, Profit.

INTRODUCTION

Agriculture is the main source of livelihood in India. Around 65% Indian population lives in the rural areas and is engaged in agriculture. Agriculture sector provides employment to near about 50% work force of India. And this is the reason that agriculture plays an important role in the economy of India. But still Indian farmers are mostly using traditional approach of agriculture which they follow for years. As with the increase of population, cultivated area of land also decreased. With the decrease of land, farmer's income also decreased. To overcome this problem everybody's focus is on increasing per unit crop yield by adopting new technologies and machinery's. Increasing crop yield helps the farmers to a certain extent, but one should also think the adverse effects of over production. In agricultural planning the main focus is to achieve maximum yield rate of crops by using advance technologies, fertilizers, seeds etc and farmers forget that beside maximizing the crop yield one should also take care the crop management techniques, so that over production of a particular crop should not happen and leads to the price loss beside having good crop yield. Researchers are using many Machine Learning (ML) algorithms in predicting the production of crop yield per unit area [1]. There were cases of loss when the yield of a particular crop is very high. In this case we can apply crop selecting method and reduce the losses. By selecting proper crop selection, losses can be converted into gain [2].

LITERATURE SURVEY

Rajneesh and Sachi [3] introduced a ML model using Softmax regression for yield prediction on the basis of various environmental factors like rain fall, soil nutrients, temperature, type of soil and various climatic conditions. Soil health card and rain data for the crops sugarcane, rice, wheat and mustard for the region of Western Uttar Pradesh in India has been used.

Pavan Patil et al. [4] proposed a system where machine learning can be used in agriculture for the betterment of farmers. The authors have used only one attribute for crop yield prediction with an aim of adding more attribute with the advancement of the project. In this paper the authors recognized several crop yield patterns for prediction. This system will be helpful in deciding that which crop should be cultivated in which particular region.

Parasuraman Kumar et al. [5] designed an IoT based smart agriculture system for automation of agriculture using artificial intelligence for making the farmers more smart and techno-friendly. This paper helps in controlling the excess use of pesticides and other soil polluting agents by using IoT devices. It also helps in using water and fertilizers in a controlled manner.

Celia M. Reyes et al. [6] suggest that agriculture insurance is the risk management tool for the farmers and how it has been implemented. In this paper the authors shares the experiences of various countries in implementing the agricultural insurance program and found that it is important for farmers yet costly to implement. The authors also found that targeting eligible beneficiaries is important in the success of agriculture insurance as it is highly subsidized.

Subhankar Mukherjee and Parthapratim Pal [7] addressed the various hurdles they found in the implementation of agricultural insurance schemes which are launched by the government of India like Pradhan Mantri Fasal Bima Yojana. The authors give suggestions to achieve the desired targets set by the governments by analysing the agricultural data and the similar schemes in the past. They also pointed out the areas where extra efforts need to done.

METHODOLOGY

In this paper a study on crop insurance has been done using the data set of sugarcane, wheat and potato by considering the strength of various calamity factors like Drought, Hail Storm, and Wind Storm etc. and the drop in yield production of the corresponding crop due to their effect. The system architecture of our proposed methodology is shown in the below figure 4.

Measuring Hailstorm Intensity

Hailstorms are the rain of solid snow ball which varies in size from very small to as big as a big stone. It can cause extensive damage to crops, plants and trees. The Torro Hailstorm Intensity Scale is used to measures and categorizes hailstorms. The scale ranges starts from 0 and goes up to 10. 0 means "no damage" and 10 means "extreme damage". In this paper our aim is to categories the loss to the crops, so we consider only the range 0 to 5 because beyond this range no crop will survive.

Table 1. Intensity of Hail Storm and its Effects

Category	Type of Damage
H0	No Damage
H1	Leaves of plants and flower petals damaged.
H2	Leaves of plants and trees are stripped.
H3	Small plants and crops are broken.
H4	Branches of small trees are broken off.
H5	Branches of large trees are broken off.
H6	Roofs are breached; metal roofs are scored; wooden window frames are broken away
H7	Roofs are shattered to expose rafters; cars are seriously damaged
H8	Shingle and tile roofs are destroyed; small tree trunks are split; people are seriously injured
H9	Concrete roofs are broken; large tree knocked down; people are at risk of fatal injuries
H10	Brick houses are damaged; people are at risk of fatal injuries

Measuring Drought

Drought can not be defined using a single factor. There are multiple factors which indicates drought, like rainfall, streamflow, snowpack and many more. We have to track these factors to monitor drought. No single factor can define the severity of drought. Even its very difficult to identify the beginning and end of drought. There are a number of tools which helps define drought [8].

The most common and widely used one such tool is The Palmer Drought Severity Index (PDSI). A drought index value is typically a single number, which is interpreted on a scale of extremely wet, average wet, and extremely dry. In PDSI researchers categories different levels of wetness and dryness that are significant in that area. The PDSI is then studied based on the data of precipitation and temperature data. The local Available Water Content (AWC) of the soil is also considered [9].

Table 2. PDSI Classification

Intensity	Effect
4.0 or More	Extremely Wet
3.0 to 3.99	Very Wet
2.0 to 2.99	Moderately Wet
1.0 to 1.99	Slightly Wet
0.5 to 0.99	Incipient Wet Spell
0.49 to -0.49	Near Normal
-0.5 to -0.99	Incipient Dry Spell
-1.0 to -1.99	Mild Drought
-2.0 to -2.99	Moderate Drought
-3.0 to -3.99	Severe Drought
-4.0 0r Less	Extreme Drought

Measuring Tropical Cyclones

The wind speed and its impacts have been defined using the Saffir-Simpson Hurricane Wind Scale. The Saffir-Simpson Hurricane Wind Scale measures the intensity of hurricane which starts from a speed of 74 mph. Wind speeds up to 38 mph are called as tropical depressions and the wind speeds from 39 mph to 73 mph are called as tropical storms.

Table 3. Saffir-Simpson Wind Scale

Wind Speed	Level of Damage
Wind (mph): 74 - 95	Minimal damage, trees can be uprooted and flooding may be in coastal areas.
Wind (mph): 96 - 110	Moderate damage, trees can be uprooted and flooding can be in coastal areas.
Wind (mph): 111 - 129	Extensive damage and serious coastal flooding to those on low lying land.
Wind (mph): 130 - 156	Extreme damage, trees blown down. Flat land may become flooded.
Wind (mph): Greater than 156	Catastrophic damage, all trees blown down.

Data Collection

The dataset of the model includes the average crop yield of sugarcane, wheat and potato of the selected area. The dataset has been taken from the governments websites during the normal cultivation conditions. The dataset also consist of attributes like different types of natural calamities and their strength that has different adverse effect on the yield results of selected crops. Further the dataset is divided in a ratio of 70\% for training and 30\% for testing. Further some part of the training data set is used for validation purpose also.

The crops are chosen as per the strength with which they can tolerate the effects of natural calamities. One strong crop like sugarcane, one moderate crop like wheat and one weak crop like potato has been chosen for each category.

Now, as most of the data is categorical data, so for using this data in our model we need to convert these categorical data into numerical data. So we convert this data to numeric data by assign a number for each category.

Table 4. Loss Due To Hail Storm

Crop	Calamity	Loss % in Yield
Sugarcane	0	0
	1	10
	2	25
	3	45
	4	60
	5	80
Wheat	0	10
	1	20
	2	40
	3	60
	4	80
	5	100
Potato	0	0
	1	20
	2	40
	3	70
	4	90
	5	100

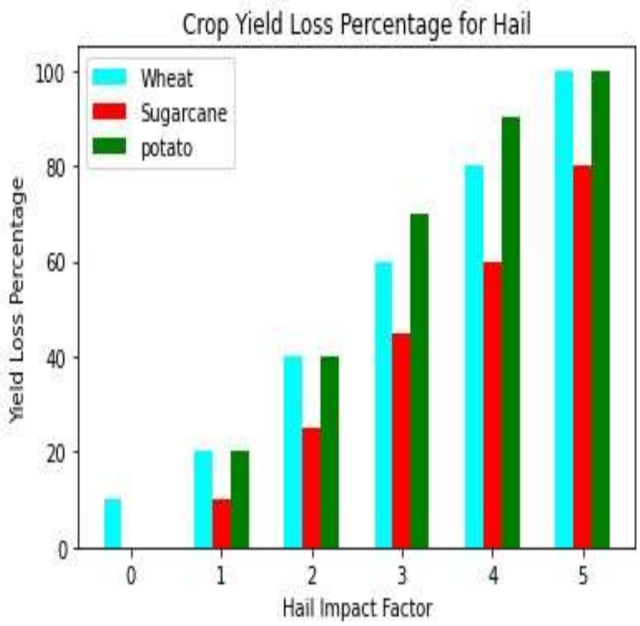

Figure 1. Crop Yield Loss Percentage for Hail Strom

Table 5. Loss Due To Drought

Crop	Calamity	Loss % in Yield
Sugarcane	1	5
	2	10
	3	25
	4	50
	5	80
Wheat	1	10
	2	25
	3	40
	4	70
	5	100
Potato	1	20
	2	40
	3	60
	4	80
	5	100

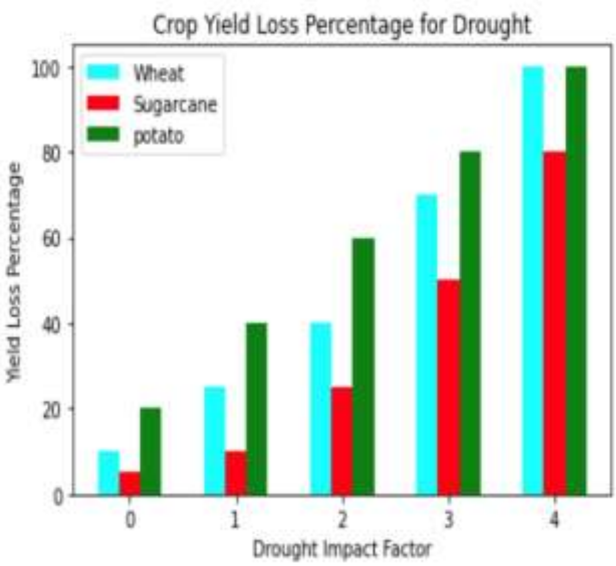

Figure 2. Crop Yield Loss Percentage for Drought

Table 6. Loss Due To Wind Strom

Crop	Calamity	Loss % in Yield
Sugarcane	1	0
	2	5
	3	10
	4	20
	5	30
Wheat	1	10
	2	20
	3	30
	4	40
	5	50

Crop	Calamity	Loss % in Yield
Potato	1	0
	2	5
	3	10
	4	15
	5	20

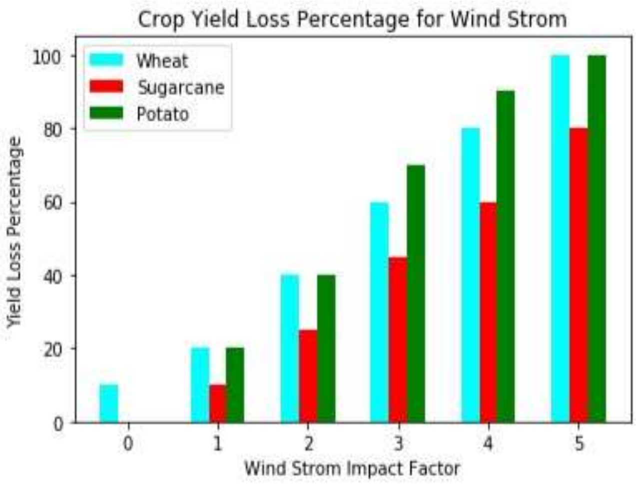

Figure 3. Crop Yield Loss Percentage for Wind Strom

Preprocessing of Raw Datasets

Data preprocessing is a part of data preparation which helps in transforming the data into a format that is more easily and effectively processed, it consist of the process of examining the data and deleting the unwanted data from the dataset. It also involves normalization of some values and replacing some null values with suitable values. After these steps the raw dataset is ready for processing.

Applying Algorithms

Various Machine Learning algorithms have been applied and tested for the selected crops data sets to calculate the loss in the yield depending upon the natural calamity and type of crop but at the end we find that Random Forest Regression best suits the given datasets.

Implementation

Implementation of the project has been done using python. In this paper we have build a model which needs average crop yield during normal climatic conditions and the strength of the natural calamity, then it calculates the loss of yield due to a particular natural calamity. The model is trained using the random forest algorithm and the datasets.

Training of the model involves following steps:

1. The very first step is to import the required libraries.
2. Upload the appropriate dataset and select the desired feature.
3. Insert the missing data.
4. Convert the categorical data into some numerical codes.
5. Separate the dataset into the ratio 70\% for Training set and 30\% for Testing set.
6. Performed the data pre-processing like Feature Scaling.
7. Select the Random Forest Regressor for the Training set.
8. Predicts the results.

Random Forest Algorithm has been used for the training purpose of the model. Steps involved in random forest algorithm are as given below:

1. Randomly select "n" number of features from total "m" number of features. (Where n $<$ m)
2. Among the "n" features, calculate the node "d" using the best split point.
3. Split the node into child nodes using the best split.
4. Steps 1 to 3 should be repeated until "l" number of nodes has been reached.
5. Repeat steps 1 to 4 for building forest for "k" number times to create "k" number of trees.

The model is trained by including more than 150 decision trees for constructing the random forest. For calculating the accuracy or performance of the model k-fold validation technique have been used. The accuracy of the model was found out to be 91% using 10-fold cross validation. Cross-validation is a statistical method used to estimate the performance (or accuracy) of machine learning models. The model needs the strength of the natural calamities during the life cycle of the crop, to make a new prediction.

System Architecture

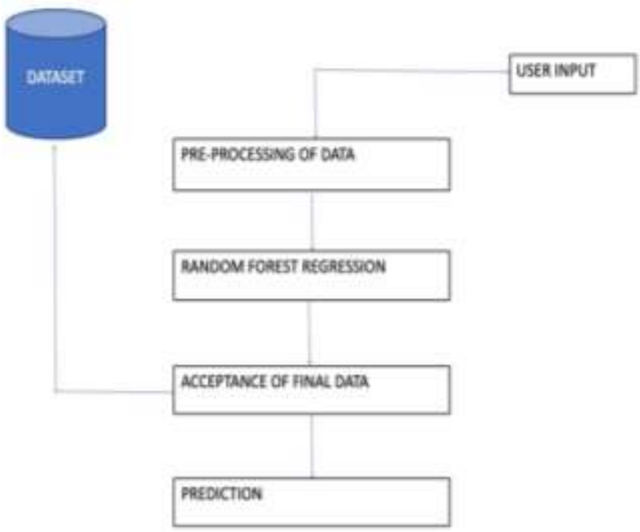

Figure 4. Shows the Architecture Diagram of the Model

EXPERIMENT AND RESULT

The aim of the proposed project is to provide an accurate procedure which gives an estimation of the loss in crop yield per unit area depending on the strength of the natural calamity using Random Forest Regression. This crop loss can then be utilized by the various insurance agencies including government agencies for calculating the loss suffered by the farmers due to these natural calamities and using this they can calculate the exact amount to be given as compensation to the farmers.

Author tested the system for various climatic conditions and their strength's data set and its adverse effects on the crop yield production during the life cycle of the crop.

CONCLUSION

In this project author shows a potential use of Machine Learning to decide the agriculture insurance amount based on the crop yield depending upon the previous year's history of crop yield production during a natural calamity and their respective prices during those years for the reason of Uttar Pradesh (West). The research has taken various calamity factors in consideration like draught, flood, hail rain and wind storm datasets of the selected area. Researcher applies many Machine Learning Models and found that Random Forest gives the suitable results for predicting the crop yields and their corresponding agriculture insurance amount. The proposed model is supported by well refined datasets.

FUTURE SCOPE

In future, we can include more crops. Along with natural calamities the model can be extended for crop deceases also. We also try to consider other region and the influential factors of that particular region. Inclusion of larger data sets of the selected area is also our concern so that the model can act more accurately, efficiently and reliable.

REFERENCES

[1] Abhijeet Pandhe et al. "Crop yield prediction based on climatic parameters". In: METHODOLOGY 6.03 (2019).
[2] V S Vyas and Surjit Singh. "Crop insurance in India: Scope for improvement". In: Economic and Political Weekly (2006), pp. 4585–4594.
[3] Rajneesh Kumar and Sachi Pandey. "An accurate prediction of crop yield using hybrid deep capsule auto encoder with softmax regression". In: Multimedia Tools and Applications (2022), pp. 1–23.
[4] Pavan Patil, Virendra Panpatil, and Shrikant Kokate. "Crop prediction system using machine learning algorithms". In: Int. Res. J. Eng. Technol. (IRJET) 7.02 (2020).
[5] Kumar Parasuraman, Udayakumar Anandan, and Anbarasakumar Anbarasan. "IoT based smart agriculture automation in artificial intelligence". In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE. 2021, pp. 420–427.
[6] Celia M. Reyes et al. Agricultural insurance program: Lessons from different country experiences. eng. PIDS Discussion Paper Series 2017-02. Quezon City, 2017.
[7] Subhankar Mukherjee and Parthapratim Pal. "Impediments to the spread of crop insurance in India". In: (2017).
[8] C T Dhanya and D Nagesh Kumar. "Data mining for evolution of association rules for droughts and floods in India using climate inputs". In: Journal of Geophysical Research: Atmospheres 114.D2 (2009).
[9] Rajneesh Kumar and Rajendra Prasad Mahapatra. "Severity of natural calamities and crop yield prediction using hybrid deep learning model in Uttar Pradesh". In: World Water Policy (2024), pp. 224 – 279.