CROWD SOURCE BASED AGRICULTURE SCHEME FOR BIG DATA ANALYTICS

D.BADRINARAYANAN[1], R.NETHRA[2], P.MALAVIKA[3], R.SRIDEVI[4]

[1]ASSISTANT PROFESSOR; DEPARTMENT OF ELECTRONIC AND COMMUNICATION ENGINEERING, PANIMALAR INSTITUTE OF TECHNOLOGY, CHENNAI.

[2-4]UG SCHOLAR, DEPARTMENT OF ELECTRONIC AND COMMUNICATION ENGINEERING, PANIMALAR INSTITUTE OF TECHNOLOGY, CHENNAI.

---***--

ABSTRACT: Agriculture, in particular in the developed world areas like India, is the best utility region. The use of record age in agriculture can replace decision-making circumstances. Farmers can produce greater yields. The lack of proper decision-making by farmers about yield prediction provides a valid explanation for this. There is no context in place to indicate what plants farmers are cultivating. As an agricultural country, India's economy is primarily dependent on growth in agriculture and agro-industry products. Data mining is an emerging field of study for the study of crop yield. In agriculture, yield prediction is a major problem. Any farmer would like to hear how much he expects to yield. We all know in India that agriculture is the country's backbone. This paper forecasts the yield of almost every crop planted in India. In the CNN (Convolutional Neural Network) plant yield prediction model, the system provides the end-user with adequate recommendations for the needed fertilizer ratios based on soil and atmospheric parameters on land, thus enhancing crop yield and increasing farmers' incomes.

KEYWORDS: crop yield forecast, demand-based crops, methods of machine learning, random forests, polynomial regression, decision tree; Learning monitored

INTRODUCTION

Agriculture is what people have begun to get up moderately, regardless of what keeps us alive for miles. However, some exciting ranchers whose lives continue to grow are moving forward. Nevertheless, the pollution which extends packages these days is also present. Our primary intention is to cost the cultivation network, which is driven behind the present focused show situation and the mission of achieving reasonable costs, a reasonable cost by making existing proof and principles solid and gradually convincing through the implementation of innovations and systems. The key incentive for creating a managed market is to postpone the unwanted exchanges, reduce the costs in the shopping centre and provide the farmers with fair costs. There were a few actions to promote rural areas with a good way of cultivating and maintaining the place of monetary improvements in the country. Besides, the inland rural promotion system in the United States of America wants to be integrated and enhanced to benefit cultivation from the new world market.

Interestingly, the contraction of the commercial centre should be rekindled in: Give Farmer more impetus to deliver. Pass on the buyers will to the buyers to enable them to arrange the production. Encourage a significant number of market players to genuine challenge and improve Farmers' offer at the last cost of their agricultural products. Currently, farmers are producing crops depending on previous age experience. Since the customary cultivation method is polished, there is a massive or scarce yield without the actual need. The farmers do not know what the contemporary horticultural economy's interest is. This leads to the farmers' misfortune. In response to suicide, low costs of production, weight and flowering of relative liabilities, the bad water system and the explosion in development costs were the communicated thinking process which was important. The main reason for this is the small costs and increased development costs of the objects. Economic interest and the containment of the product controls the costs of yield. The yield forecast for the bleeding edge ML calculations is one of the companies that should be likely. Field sensors, satellites, UAVs and farm equipment may provide records about soil conditions, physiology of plants, the environment and several local operations at a house.

These data sets authorize the method and estimation forms which could be very useful for agricultural production. The call for a vertical improvement in maintenance safety in agriculture is high and could blast along those lines in the coming predetermination. A combined fundamental and methodological approach includes starting assortment, pesticides and composting of managers, consolidated editing and water collection, competent water system technology and many more. It

also becomes essential to reproduce and anticipate the harvest yields under circumstances previous use to ensure viable yield for managers and preferred results, more remarkable in sprinkled areas and India, when researchers cultivate in closer proximity to the exactness. Since the relative values between harvest yield and climatic and non-atmospheric factors are not straightforward and present reasonable dimension problems, the machine examination will show an optional victory over output expectations.

SCOPE OF THE PROJECT

The objective of the project is to ascertain the crop yield of a region, by taking into consideration the data set containing certain significant or crop-related characteristics such as temperature, moisture, precipitation and crop production in previous years. Regression models are used to predict a continuous value. It's a procedure monitored. During the training and development, the regression model is pre-processed and fits into the trained data. Here, the emphasis is on reducing cost by finding the best fit. The output function makes error measurement easier. During training, errors between the forecast and the real values are decreased to minimize the error function.

OBJECTIVE

The farmers focused on how agriculture can contribute to economic growth and modernization as a whole. Robust farm growth and productivity are deemed to be crucial to sustainable economic development. In developed nations, there are many policymakers, analysts and academics. In agriculture, the management of pests and diseases for cultivated crops is one of the major issues facing farmers. Usually, production costs and high health risks rise significantly because of excessive use of pesticides, due largely to lack of information on diseases and pests in farms.

CHALLENGES

● Challenges are a big factor that will have a negative impact on the current project.

● Choosing an appropriate dataset, then tuning the parameters to make the project more effective to get the desired results are some of the problems encountered during crop yield prediction.

● The model must be equipped with less computing efficiency and power in mind.

● Due to the dynamically changing climate, the error rate has increased.

LITERATURE REVIEW

In [1] TITLE-Crop Yield Prediction based on

Indian Agriculture using Machine Learning. **AUTHOR-**Potnuru Sai Nishant, Pinapa SaiVenkat, Bollu Lakshmi Avinash,

B.Jabber. **DESCRIPTION-**The paper uses advanced regression techniques like Kernel Ridge,Lasso and ENet algorithms to predict the yield

In [2] TITLE-Design and Implementation of crop yield prediction model in agriculture. **AUTHOR-**Sangeetha, Shruthi. **DESCRIPTION-**In this work, they used machine learning approach by analyzing various factors like rainfall, temperature, humidity, soil, Ph ,soil type, etc.

In[3] TITLE-Predicting Yield of the crop using Machine Learning Algorithm. **AUTHOR-**P. Priya, U.Muthaiah, M.Balamurugan. **DESCRIPTION-** This paper predicts the yield based on existing data by using Random Forest Algorithm .Real data of Tamil Nadu were used for building models

In[4]TITLE-Crop Yield Prediction Using Machine Learning. **AUTHOR–**Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara, Mansing Rathod**. DESCRIPTION-** This paper predicts the crop yield can be used by any user their choice of crop.

EXISTING SYSTEM

In the current system, the output of the crop depends on the seasonal environment. Climate conditions in India differ unreservedly. Farmers face severe challenges in times of famine. So they used those algorithms for machine learning to help farmers propose the crop to improve yield. In order to estimate future information, they take several data from previous years. In WEKA, they classified the findings using SMO classifiers. The major considerations were minimum temperature, maximum temperature, average temperature, and information on crops and yields from previous years. They have divided the past data into two as high yield and low yield by means of SMO.

DRAWBACKS

● Lack of experience with formers

● Changes in the climate at a rapid pace

● Increase in the Number of Earth Disorder

● There is no communication between the formers and the Technical Team.

METHODOLOGY

Machine learning primarily consists of three learning techniques for training a model: supervised learning, reinforcement learning, and unsupervised learning. Supervised learning is a method of learning that maps are known inputs into outputs that are mapped from input to output. However, in the case of unsupervised learning, we would not know the desired output. In this case, we would need to train the model to get the desired output.

DESCRIPTION OF THE DATASET

For crop yield prediction, most researchers used.csv files from agricultural datasets. It's a supervised learning dataset. It has many characteristics such as County Name, State, Humidity ,Temperature, NDVI, Wind Speed, Yield, and so on.

PROPOSED SYSTEM

We improve the Prediction of crop yield using an effective algorithm in the proposed system. The challenge is to create the most effective model to predict the crop's output, so experiment with various algorithms and compare them to see which one has the least error and loss, then choose that model and predict the crop's yield. We use a real-time price forecast system, so the outcome will be dynamic and change depending on the day's events.

ADVANTAGES

● The project's ultimate goal is to produce the most crop with the lowest yield.

● Farmers should benefit from early detection and management of issues in order to increase crop yield.

● To gain a better understanding of crop yield, we must analyze large amounts of data provide an accurate yield for that crop and advise the farmer on how to grow a better crop.

COLLECTION OF INPUT DATA

Data can come from a variety of places and must be compiled before it can be used. Our information came from a variety of government and private websites. The table below contains a list of all of them.

In the case of rainfall data, the above-mentioned website provided directly importable data. However, historical data on pressure and temperature was available, but it had not been gathered. We used Python's Beautiful Soup package to compile this information. Beautiful Soup is a Python library for parsing HTML and XML files and extracting data.

DATASOURCE:

http://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-list-raingauge-stations-india-meteorological-department **- For rainfall**

DATASOURCE: http://www.timeanddate.com–**For temperature and Pressure**

DATASOURCE:

http://mahaagri.gov.in/cropwatch/asp/rpt1.asp   **-For crop statistics**

DATASOURCE:

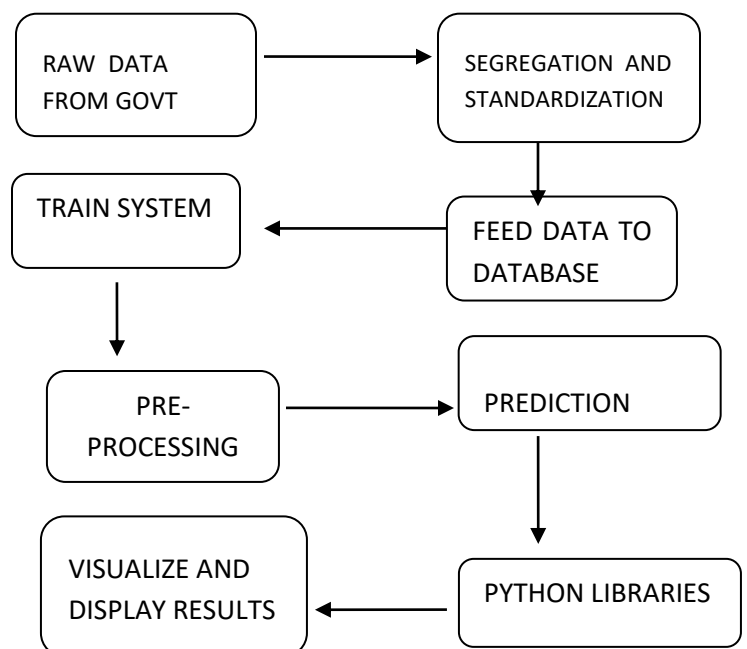http://farmer.gov.in/Drought/DroughtExport.as px?5+z9

6qEraJI =_- **Government**

Announced droughts through

Tamil Nadu

ARCHITECTURAL DESIGN:

The system's flow has been designed as follows:

● The raw data is gathered.

● The information is separated and standardized.

● It is entered into the database; we will use Excel as our database.

● It is then sent for pre-processing, after which the data is fed into the system to train it.

● Algorithms are used to forecast outcomes.

● After that, Python scripts and libraries would be used. Visualize the findings to get a better sense of the trends.

SYSTEM REQUIREMENTS

● **HARDWARE REQUIREMENTS:**

● System: Pentium Dual-Core.

● Hard Disk: 120 GB.

● Monitor: 15'' LED

● Input Devices: Keyboard, Mouse

● Ram: 1 GB

● **SOFTWARE REQUIREMENTS:**

● Operating system: Windows 7.

● Coding Language: Python

● Database: MYSQL

RESULTS AND DISCUSSION

Machine learning frameworks were used extensively in this study. The experiment was conducted on a DELL personal laptop with an Intel Core i3-4300U CPU running at 1.90 GHz and 8 GB of memory. This section looks at the significance of the project's outcomes. We've picked two dataset files for this project, all of which are in. The data we considered for the experiment is in the CSV format, which has a total size of 5l.9MB. 70% of the data will be used for instruction, and 30% will be used for tests. Initially, we select samples for model training and choose some of the characteristics such as yield, temperature, and location. Following the successful training and testing of the dataset, we moved on to determining the model's accuracy. The accuracy of the model indicates how well we predicted the crop yield in comparison to the original data;

the closer the model's accuracy is to the original yield value, the better. The yield variable is plotted with various characteristics such as state, pressure, temperature max, temperature min, and so on for data analysis.

CONCLUSION

The application of machine learning and data mining algorithms in agriculture is highlighted in this project. Crop productivity and drought forecasts, if provided in a clear and understandable manner to the end-users, the farmers, would be of great assistance to drought-stricken villages and districts. Overall, the Iterative Dichotomizer 3 algorithm proved to be the most precise in delivering forecasts for Classifier1 in our project, as well as the best results for the Sugarcane crop in Classifier2. Classifier1–Drought Classification had an accuracy of 83.92 per cent, while

Classifier2–Crop Productivity had an accuracy of 44.68 per cent. These findings can be improved upon and used in future studies as well.

FUTURE SCOPE

After successfully introducing it in various divisions in the state of Tamil Nadu, this system can be implemented in other states and countries in the future, solely to relieve the misery of farmers. Governments from all states facing similar difficulties may use this system to find solutions.

REFERENCES

1. Abd Rahman, A., 1992, ‗Agricultural Development in Malaysia: Retrospect and Prospects, Journal of Economic Cooperation among Islamic Countries, 13:3-4: pp 51-86.

2. Wong, L. C. Y., 2006, ‗Implications of Regional and Global Developments on Agriculture and Its Competitiveness in CLMV: Increasing Importance of Supply Chains and International Trading Networks‗, paper presented at SEARCA Policy Round Table on ‗ Shoring and Sharing Capabilities to Improve Agriculture and its Competitiveness in Transition Southeast Asia‗, Siem Reap, Cambodia, 24-25 April 2006.

3. FAO, 2008a. The state of food insecurity in the world: High food prices and food security– threats and opportunities. http://www.fao.org/SOF/sofi/.

4. NKEA-Agriculture@pemandu.gov.my.

5. Abdullah Chek Sahamat, Native Customary Right Land Development: An Approach to Improve the Rural Native

Wellbeing of Sarawak, Seminar Kebangsaan Perancangan Bandar dan Wilayah Ke-23 (Skudai: UTM, 21 September 2005).

6.http://etp.pemandu.gov.my/annualreport2012/upload/ Eng_ETP2012_07_NKEA05_POR.pdf.

7.http://www.pbs.org/newshour/rundown/2013/05/in-south-africa-using-mobile-technology-To improve-maternalhealth-access.html.