A Predictive Modeling Framework for Enhancing Crop Yield using Automated Data Driven Approach

Devidas[1], Balasubramani.R[2], Jason Martis[3]
[1]Assistant Professor ,Dept. of ISE,NMAM Institute of Technology, Nitte, India, devidasbhat@nitte.edu.in
[2]Professor, Dept. of ISE,NMAM Institute of Technology, Nitte, India, balasubramani.r@nitte.edu.in
[3]Assistant Professor ,Dept. of ISE, NMAM Institute of Technology, Nitte, India, jason1987martis@nitte.edu.in

ABSTRACT

Over recent years, there is a surplus mortality rate among the farmers. The reason behind this cause includes the weather condition, debt, changes in the policies in the loan repayments. In general, the farmers lack the awareness in identifying the type of crop which suits the soil quality, its nutrients and the soil composition. Our paper focuses on addressing farmers issues by improving crop and soil quality. We accomplish this setup by performing an extensive analysis of soil support for crops. Our proposed system overall achieved a better prediction scenario of 80%, which is comparatively more significant than existing systems.

Key words: *Data Mining, Soil Quality, Crop predictions, Machine Learning, Agriculture*

1. INTRODUCTION

For a country, the most crucial aspect of its development is its capacity to produce more food. For decades, agriculture was a leading source for the production of essential crops. The rate of urbanization at present is by-far the most prime aim of our civilization. As the amount of land will not be increasing in this era of urbanization, we have to device newer ways to farm arable lands and extract the absolute most from these limited land resources. In this age of technology, with the proper use of data science, the agricultural sector can be changed for the better. Farmers alone are indeed the best deciders of crop cultivation and selection. However, numerous machine learning techniques are utilized in this field for better precision and stability of results. The main thing Indian farmers lack is the knowledge of soil and suitable techniques used to improve soil quality. Foreign agriculture industries heavily rely on machines as well as perform micromanagement on soil assessment, thus improving crop yields. However, Indian farmers are not well educated and depend on ancient methods of cultivation. This bridges a considerable gap in soil and crop productivity.

Machine learning and data mining are extensively in use in other fields to mine necessary information required for stakeholders. Researchers have also claimed that machines have become much better at handling day to day mundane activities. Weather prediction is one such setting which can predict even impending storms way before they form, thus saving millions of lives and property. There is also the integration of machines in day to day lives of people, thus improving human quality of life. Similarly, we can also integrate machine intelligence in the farmer's life. These systems can assist farmers with understanding the significance of harvest expectation, with building their necessary information about soil quality, and by understanding the area savvy climate restrictions, and to have an expansion crop yield through innovative and soil friendly techniques. The more significant part of existing frameworks are equipment related which makes them exorbitant and is hard to keep up. Likewise, they even need to give exact outcomes. A few frameworks recommend the harvest grouping relying on the yield and market cost. These frameworks attempt to conquer this kind of disadvantages by detailed prediction the harvests by breaking down the organized information, expecting soil quality utilizing information mining approach for the most part center around the horticultural viewpoints.

2. RELATED WORK

The Agriculture division assumes an imperative job in the Indian Economy. Technology usage in the field of agriculture is very minimal in India, especially in the field of agriculture. Due to lack of technical support and less usage of these advancements, there is a severe decrease in the farming's commitment to the Gross Value Added. Nourishment one of an essential thing for life and everybody relies upon the agrarian outcome. The ranchers assume a significant job. The accompanying examination are as below

Investigations in the article [1] reveal that the authors have used the Multiple Linear Regression technique for the yield analysis. The Decision tree calculation and Classification is executed to play out the examination of the complete 362 data sets and give the outcome. The available preparation data-set is delegated in inorganic, natural and land for the forecasting soil conditions. Results given by this framework are exceptionally exact and reliable.

Article [2] stretches information on the practice of a Back Propagation Network (BPN) which assesses the test data-set. Back Propagation Network (BPN) utilizes a kind of concealed layer which helps the better execution for anticipating the sort of soil properties. This network is utilized to build up a kind of self-prepared capacity to anticipate the sort of soil properties with parameters provides exactness. It likewise performs route superior to conventional techniques. In any case, a portion of the occasions the framework will, in general, become too moderate and conflicting.

In a research article [3], the authors use two relapses regulated AI strategies are utilized, Support Vector Machine (SVM) along with Relevance Vector Machine (RVM) used to show the exactness in soil quality forecast. A remote keen gadget is utilized for detecting soil dampness and meteorological information. The remote gadget will give the mistake pace of 15 per cent and 95% precision. Be that as it may, it has not been utilized in the testing for the ongoing information.

Article [4] takes in the advantage of Back Propagation method for the assessment of the Soil Fertility along with Plant Nutrient. The outcomes are exact, and it empowers the improvement of the dirt properties. It performs way better when contrasted with the conventional kind of techniques. In any case, the framework is exceptionally moderate, wasteful and not steady.

Article [5] shows the different techniques in machine learning which are utilized and incorporates the Decision tree method, Naive Bayes Classifiers, and the KNN Classifiers will investigate the dirt and foresee yield of harvest. Though the Rule-Based Induction and SVM can be utilised for increment the exactness as results given are not precise. Paper 6 Discusses different approaches that can be used to plot data using different approaches using tools such as GIS-Spark.

Article [6] embellishes on a technique known as Crop Selection Method (CSM), which can be applied to solve this crop selection problem. It, in turn, increase the net gain rate of agricultural yield over the different season and thereby contribute to the economic growth of the country in a better and precise way. Crop selection technique uses a method for selecting crop(s) based on a specific season depending on the different environmental and economic factors for the maximum profitability. There are various factors such as precipitation levels, mean temperature, soil characteristic, market prices and demand, prevailing farm conditions, crop or varietal adaptability, resistance to pests and diseases, farming system, available technology and many more. The right choice in the selection of crop or crops to be grown, mainly perennial types, will eventually convert into a successful cultivating venture.

Machine learning is the branch of computer science which is used to build algorithms which exhibit self-learning property. It is viewed as one of the significant areas under Artificial Intelligence. To show intelligence machine needs to interpret and analyze the input after analyzing it the result data apart from merely following the instructions on that data. This is the thing that machine learning algorithms do. Machine learning focuses on the development of computer algorithms which can get data and make them learn on themselves based on their learning. The way toward learning starts with perceptions on information, for example, direct experiences or set of rules can be used in order to look for patterns in data. It helps to make better analysis and predictions in the future based on the samples that we give. The essential point is to make the computer system to learn automatically with minimal human intercession or help and regulate actions consequently. It is a significant field in computer science which is being utilized in various forms of progressive technological development programs all around the world. It is regarded as the future of engineering and Artificial Intelligence.

3. PROPOSED SYSTEM

This paper aims to propose and implements a rule-based model to predict accurate crop yield production based on the collection of historical data. We achieve this by expansively collecting the real-time data and comparing it against the past collected data. Our system utilizes the amalgam of several machine learning approaches to predict crop yield patterns and forecasts the appropriate crop by identifying different soil parameters on local atmospheric conditions. Fig 1. shows the overall framework of the system.
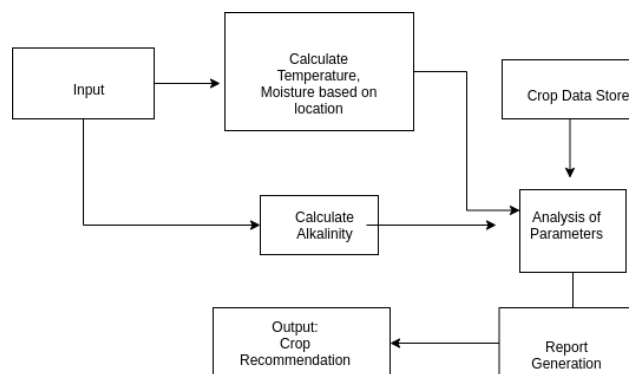

Figure 1: Figure depicting the model framework for the proposed system.

Our framework is structured in such a way to enable small scale as well as large scale businesses to develop and develop best practices for better yield generation. It also assists in achieving the degree of exactness of anticipating the harvests. The system examines the supplements available in the soil and predicts the yield profitability dependent on the area where the farmer resides.
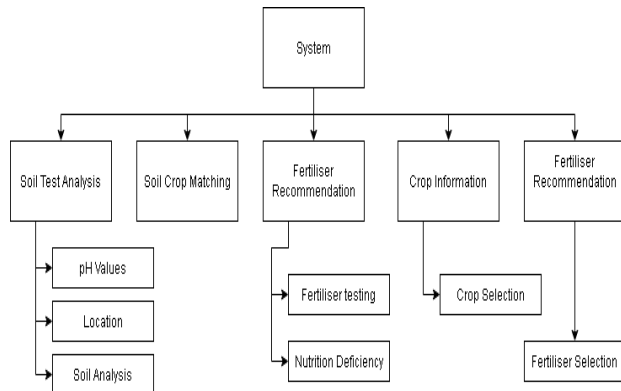
Figure 2: Figure depicting the outcomes of the proposed system.

Our framework has got five modules, as appeared in Fig 1. In soil test investigation, the client enters the pH esteem and the area. The yield of this module will be the examination result where the level of supplements in that dirt is gotten.

Soil crop coordinating module will locate the coordinating yield that can be developed in that dirt by contrasting and the harvest database. In the compost proposal module, the client will be prescribed with the manure that will give the highest harvest yield.

In the yield data module, the client can choose a harvest and can see the data about the harvest. In the compost data module, the client can choose manure and afterwards show the data about it.

It is observed that by utilizing the solo and directed learning calculations, such as Multiple Linear Regression and Back Propagation Network. The outcome of this aggregated from various system learning methods, and the most precise outcome is set to the end client. Alongside this, the end client is given a reasonable proposal about the nourishment's appropriate for every other yield. Ruling data of various crops are composed over all seasons. The framework will assess soil quality and foresee the pace of harvest yield. It likewise gives compost suggestion on the off chance that the soil requires it.

The convenience of configuration is according to the user's flexibility. The system takes the wellsprings of data that is pH (considering the level of the enhancement) and the zone from the customer. The two controllers do result taking care of region and is used as a commitment to each other. We take advantage of this process this by the usage of the pariah applications like API for the atmosphere and temperature, soil type, supplement estimation of the soil in that particular region, proportion of precipitation in that district, soil blend and many more.

The modules in the framework work in tandem with each other determining data from each other but working independently. For example, module 1 determines only environmental conditions of local soil by agglomerating past data a well as current soil conditions. Module 2 determines soil alkalinity of the earth in local conditions.

Even soil nutritional levels like Manganese, Nitrogen (N), Potassium (K), Sulfur (S), Phosphorous (P), Magnesium (Mg), Boron (B), Calcium (Ca), Iron (Fe), Zinc (Zn) are also analyzed.

The outcome of Module 1 and Module 2 are discriminated, and the predefined "supplements" data store is checked. The differentiated results are used in the third module where the blends of the above-obtained results and a predefined instructive file are accessed. Finally, the results are showing up as visual diagrams near the precision rate.

4. IMPLEMENTATION

For training the data-set, seasonal data-set including the crop information were used provided by the government. We conjure our data-set on a specific region as the model uses location-based details. The system was run on Dual Intel Xeon ES-2609V4 8C 2.0Ghz with 128GB of RAM . A total of hundred epochs and varying batch size were used during the training. The Implementation of this work is done using machine learning techniques on python language. Keras and Tensorflow as a back-end engine which supports the Scikit learn library. The seasonal crop area of each plant is shown in Figure 3.

Figure 3: Figure depicting the type of plants grown in two different seasons.
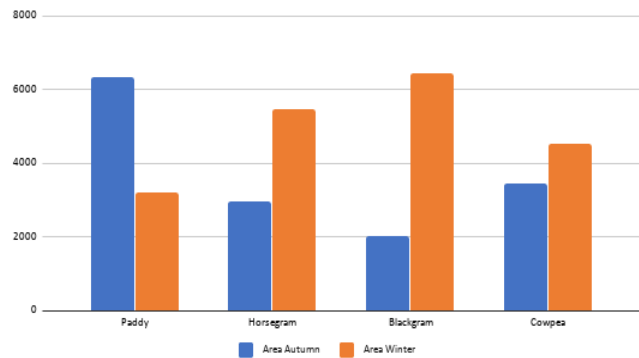
We have also performed a plot of different data obtained. Figure 4 represents the general architecture of the system.
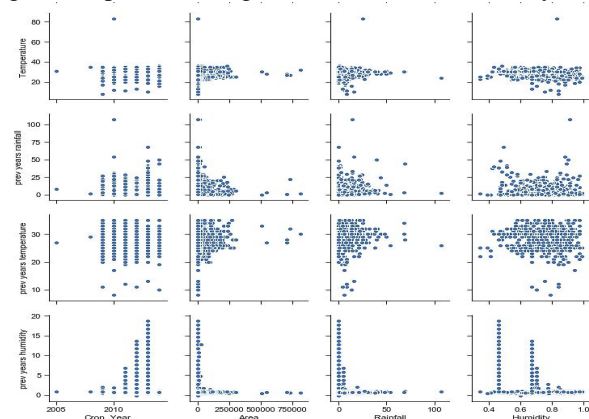

Figure 4: Figure depicting the Data Distribution Plot of various parameters.

Our paper tries to perform the analysis of machine learning methods such as SVM, Decision Trees, Random forest, Back propagation Regression are implemented, and their accuracy is plotted in graph to depict their accuracy-check. The results indicates that the Regression method yields high accuracy compared to the other methods for the sample set.
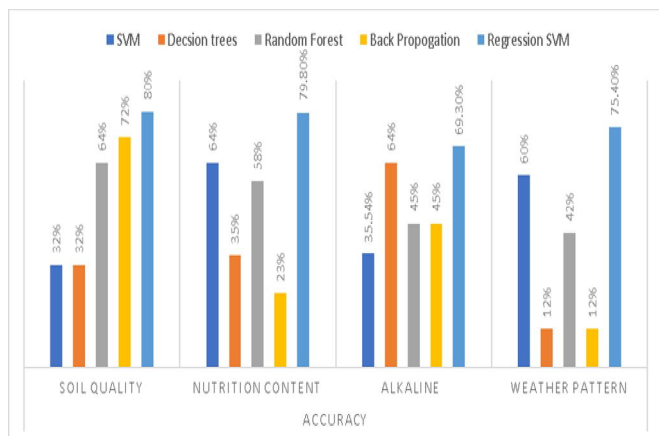

Figure 5: Figure depicting the accuracy of various machine learning techniques

From figure 5 we can come to a decision that Regressive SVM is having higher accuracy compared to the other models across all the parameters considered.

Table 1 gives a comparative study of the various machine learning techniques discussed in the paper. SVM technique gives high accuracy of 64% for Nutrition content attribute across other attributes. Decision tree method yields with an accuracy of 64% for alkaline soil property. Random Forest and Back Propagation techniques works well with soil quality with an accuracy of 64% and 72% each when compared with the attributes considered. Regression SVM gives highest overall accuracy of 76% for all the attributes considered when compared to other machine learning techniques discussed in this paper.

Table 1: Comparative Study of various machine learning techniques

	Attribute	SVM	Decision Tree	Random Forest	Back Propagation	Regression SVM
Accuracy	Soil Quality	32%	32%	64%	72%	80%
	Nutrition Content	64%	35%	58%	23%	79.80%
	Alkaline	35.54%	64%	45%	45%	69.30%
	Weather Pattern	60%	12%	42%	12%	75.40%
	Overall Accuracy	48%	36%	52%	38%	76%

5. CONCLUSION

The farmers of India use traditional methods of crop cultivation. They are unable to compete or adapt to the automation driven, and data have driven modern farming techniques. Machine learning methods are used in many fields to improve data analysis and prediction. This paper recommends a technique to embed machine learning analysis in Indian agricultural setup and predicted accuracy of environmental conditions such as alkalinity, Soil pH, weather and nutritional content. Our methods achieved higher overall accuracy compared to standard traditional approaches.

Our proposed framework utilizes administered and solo Machine learning calculations, and it gives the best outcomes, which depends on the exactness. The consequences of the given two calculations will be then analyzed, and the one which gives the best and the exact outcome will be chosen. In this manner, the framework will assist with diminishing the challenges which are looked by the ranchers and will prevent them from endeavoring suicides. It will go about as a medium to give the ranchers a proficient data which is required to get high return and in this way augment the benefits which thereby help and reduce farmers troubles due to unavailability of accurate information.

We can further improvise our framework to preserve soil quality by assisting farmers in using sustainable farming techniques.

REFERENCES

1. P. Vinciya, Dr. A. Valarmathi,**"Analysis for Next Generation High Tech Farming in Data Mining**" IJARCSSE,vol. 6, Issue 5, 2016.
2. Santanu Koley ,Shivnath Ghosh, "**Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks**" IJRITCC, vol. 2, Issue 2,292-297,2014.
3. Zhihao Hong,Z. . K. Iyer, Kalbarczyk,R "**A Data-Driven Approach to Soil Moisture Collection and Prediction**"
4. Sabri Arik, Tingwen Huang, Weng Kin Lai, Qingshan Liu , "**Soil Property Prediction: An Extreme Learning Machine Approach**" vol. 3, Issue 4,666-680,2015.
5. Devidas, Balasubramani.R ," **Groundwater Quality Mapping using Spark-GIS**",International Journal of Engineering and Advanced Technology , Volume-8 Issue-6,2019.
6. Kumar, R., Singh, M., Kumar, P., and Singh, J. (2015). **Crop selection method to maximise crop yield rate using machine learning technique. In Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials"** (ICSTM), 2015 International Conference on, pages 138–145. IEEE
7. Devidas,Balasubramani,"**Weather Prediction using Data Mining Techniques"**, International Conference in Recent Innovations in Electrical ,Electronics & Communication Engineering,(ICRIEECE)2018

8. Pradeep, B., Balasubramani, R., Martis, J. E., & Sannidhan, M. S. (2019). **Generic IoT Platform for Analytics in Agriculture**. Studies in Big Data Internet of Things and Analytics for Agriculture, Volume 2, 225-248. doi:10.1007/978-981-15-0663-5_11

9. S. K. A. S, S. M. S and A. Bhandary, "**Lemon Maturity Estimator: An Approach Using Color Image Processing Techniques**," *2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimisation Techniques (ICEECCOT)*, Msyuru, India, 2018, pp. 1213-1218, doi: 10.1109/ICEECCOT43722.2018.9001657.

10. Chen, T. J., Chou, L. F., & Hwang, S. J. (2003). **Application of a data-mining technique to analyse coprescription patterns for antacids in Taiwan.** *Clinical therapeutics*, *25*(9), 2453-2463.

11. Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). **Analysis of agriculture data using data mining techniques: application of big data**. *Journal of Big data*, *4*(1), 20.

12. Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E., & Borůvka, L. (2017). **Agricultural soil spectral response and properties assessment: effects of measurement protocol and data mining technique**. *Remote Sensing*, *9*(10), 1078.

13. Trajanov, A., Spiegel, H., Debeljak, M., & Sandén, T. (2019). **Using data mining techniques to model primary productivity from international long-term ecological research** (ILTER) agricultural experiments in Austria. *Regional Environmental Change*, *19*(2), 325-337.

14. Muangprathub, J., Boonnam, N., Kajornkasirat, S., Lekbangpong, N., Wanichsombat, A., & Nillaor, P. (2019). **IoT and agriculture data analysis for smart farm**. Computers and electronics in agriculture, 156, 467-474.

15. Anton, C. A., Matei, O., & Avram, A. (2019, April). **Collaborative data mining in agriculture for prediction of soil moisture and temperature**. In *Computer Science On-line Conference* (pp. 141-151). Springer, Cham

16. Sannidhan, M. S., Jason Elroy Martis, and Abhir Bhandary. "**A cost effective approach for detecting electricity theft using raspberry pi board**." *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE, 2017.

17. Sukhada, C., Sannidhan, M. S., Sudeepa, K. B., & Abhir, B. (2019). **A Study on various state of the art of the Art Face Recognition System using Deep Learning Techniques**. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*, 1590-1600.

18. V, S. (2019). **A High Performance on Extemporize Yield of Horticulture Crops with Predictions based Water and Soil properties using Multivariate Analytics and Machine Learning Algorithms**. *International Journal of Advanced Trends in Computer Science and Engineering, 6*(4), 1706-1710. doi:10.30534/ijatcse/2019/98842019

19. *Akshaya, Ranjan Kumar, H.S., Bhat Devidas .***A novel method to recognize object in Images using Convolution Neural Networks,2019** *International Conference on Intelligent Computing and Control Systems, ICCS 2019, pp. 425-430.*