Machine Learning Approach for Crop Yield Prediction and Crop Variety Recommendation in Android Application

Jyoti Gupta, Ayushi Chauhan, Aastha Agarwal, Abhijeet Singh Ranghuvanshi, Richa Saxena
Department of Computer Science and Engineering,
Moradabad Institute of Technology, Moradabad, U.P., India

ABSTRACT

Agriculture in India plays a major role in economy and employment. The common difficulty present among the Indian farmers are they don't opt for the proper crop based on their soil necessities. Because of this productivity is affected. This method is characterized by a soil database collected from the farm, crop provided by agricultural experts, achievement of parameters such as soil through soil testing lab dataset. The data from soil testing lab given to recommendation system it will use the collect data and do ensemble model with majority voting technique using Random Forest Repressor and Artificial Neural Network as learners to predict the crop yield and recommend a crop for site specific parameter with high accuracy and efficiency. The proposed system will integrate the data obtained from repository, weather department and by applying machine learning algorithm: a prediction of best suited crops according to current environmental conditions is made. This provides a farmer with sort of options of crops which will be cultivated. Therefore, the final result of the project that develops a system by integrating data from different types of sources, prediction analysis which may improve crop yield productivity , data analytics, and increase the ultimate profit margins of the end user helping them lifetime.

Keywords: Agriculture crop prediction, recommendation system, Neural Network, Decision Tree, Android Application.

1. Introduction

Agriculture gave birth to civilization. India is an agrarian country and its economy largely based upon crop produc- tivity. Thus agriculture is that the backbone of all business in India. Now India stands in second rank in worldwide in farm production. India is an agricultural country but remains using traditional ways of recommendations for agricultural
purpose. Presently, recommendations for farmers are supported one to at least one interaction between farmers and therefore the experts and different experts have different recommendations. The Android OS smart phones being the foremost vastly used technology are often used for recommendation of the fertilizers and crops. The Android OS is especially used within several sorts of embedded & mobile platforms, including mobile phones and tablets. and therefore the industry is investing the power of Android within other embedded platforms, that need the power to satisfy deadlines as a prerequisite for reliable operation. In our project the crop yield classification will perform to categorize on the thought of yield productivity and class labels are getting to be low, mid, and high. And range of productivity are getting to be defined and regression are getting to be performed to urge the actual crop yield estimated cost. this is often the motive to develop this technique . supported crop weather studies, crop yield forecast models are prepared for estimating yield much before actual harvest of the crops.

By use of empirical statistical models using correlation and regression technique crops yield are forecast on an operational basis for the country. Meteorological parameters at various crop growth stages in conjunction with technological trends are utilized within the models. And this research also will helpful if in future we make an entire recommender system for farmers. Because here we are performing descriptive analytics which is that the base or foundation of any recommender system.

2. Proposed System Architecture

The user has got to create an account and log into it. The random forest algorithm is employed to predict the crop yield and neural network algorithm is employed for crop recommendation to the users. The users can open their app to look at recommended crops and weather and government schemes provided to them.
This application is extremely useful in terms of its applications because it provides both recommendation also because the feature to be notified about weather outlook.

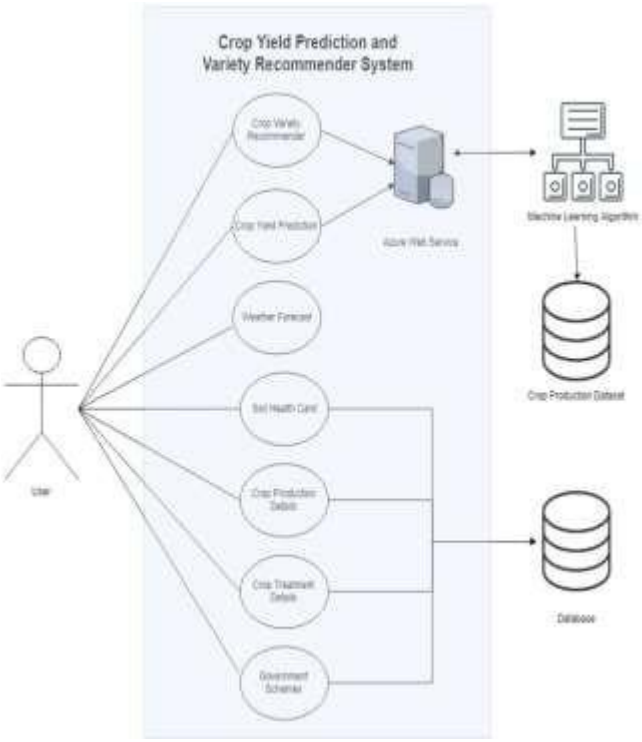

Fig: Use Case Diagram of the Proposed System

3. METHODOLOGY

The paper shows the importance of crop selection and the factors deciding the crop selection like production rate, market price and government policies are discussed. This paper proposes a Crop Selection Method (CSM) which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crop to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence thereis a need to include a prediction method with improved accuracy and performance.

Fig : System Architecture of the project

2.1 Dataset Collection

Title: District-wise, season-wise crop production statistics
Description: The data refers to district wise, crop wise, season wise and year wise data on crop covered area and production. The data is being used to study and analyse crop production, production contribution to district/State/country, Agro-climatic zone wise performance, and high yield production order for crops, crop growing pattern and diversification. The system is also a vital source for formulating crop related schemes and assessing their impacts.
Released Under: National Data Sharing and Accessibility Policy (NDSAP)**Released Under**: National Data Sharing and Accessibility Policy (NDSAP)Dataset link: https://data.gov.in/resources/district-wise-season-wise-crop-production-statistics-1997

rows 246091 columns 7

State_Name	District_Name	Crop_Year	Season	Crop	Area
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102
Andaman and Nicobar	NICOBARS	2000	Whole Year	Banana	176

2.2 Crop Variety Recommendation System Using Artificial Neural Network

The dataset for which we have taken the data as a training set and tried applying the algorithms on it by taking the data of past as a test set and then view the output. This obtained output is compared with the actual output. Crop with maximum points are often recommended to the farmer. The market trend of the crops is saved in the database. While recommending quite one among the crops, the primary factor determined are going to be the year factor which will be followed by market factor and therefore the ratio factor. For recommending the crop to the user, we are using the Multi-Class Neural Network. In Neural Networks some nodes use a nonlinear activation operate that was developed to model the frequency of action potentials, or firing, of biological neurons[4].

Input: The prediction of crop is dependent on numerous factors such as Soil type, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system.

Data Acquisition: counting on the present user location, the system mines the soil properties within the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set.

Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multi- class neural network is used by the system to predict the crop type. The prediction is based on past production data of crops i.e.: identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner.

Output: The most profitable crop is predicted by the system using Multi class neural network algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

Set Theory:

S = {I, Fm, O, S, F}
I = {I1} set of Input.
I1 = Location of user
Fm = { GetLocation(), GetAttributes(latitude, longitude), GetSoil() , GetWeather(), FeasibleCrop(soil, weather), PastProduction(FeasibleCrops)} Set of functions.
Weather – Temperature and Rainfall values
O = {Crop predicted for given Location} Set of output.
S= Correct prediction for High production and profit Success Condition
F = Failure in prediction due to incorrect training data Failure Condition

Train data:

$Y1 = 0 + \beta 1 xi2 + \ldots\ldots + \beta p\ xip + \in i$ for i=1,2, …, n Where, $\beta 0$, $\beta 1$, $\beta 2$…are coefficients of Multi class neural network
$xi1, xi2 \ldots\ldots\ldots\ldots xip$ are independent variables.
X {weather attributes, soil attributes}
Y{production}
$Y = X\beta + E$ Y = Xβ + E
Y- production matrix
X- attributes matrix
B- Partial coefficient matrix

E- error control
β = (X'X)-1 X'Y Least Square Estimate
X' - Transpose X-1 - Inverse of Matrix
Prediction: Y = X β
Result: Y-Y'

Table 1.0. Prediction accuracy of developed model for different crops.

S. No. Name of the Crop Average prediction accuracy, %
1 Soyabean 92
2 Paddy 88
3 Maize 91
4 Wheat 90
E- error control
β = (X'X)-1 X'Y Least Square Estimate
X' - Transpose X-1 - Inverse of Matrix
Prediction: Y = X β
Result: Y-Y'

Table 1.0. Prediction accuracy of developed model for different crops.

S. No.	Name of the Crop	Average prediction accuracy, %
1	Soyabean	92
2	Paddy	88
3	Maize	91
4	Wheat	90

The prediction accuracy of the developed model varied from 88 to 93 per cent for the selected crops and selected districts. Based on these observations the overall prediction accuracy of the developed model is 91.00 per cent. With a high prediction accuracy the developed model can be used by the policy makers in arriving at a policy decision well in advance i.e., before the harvest of the crop.

2.3 Crop Yield Prediction System Using Random Forest Regressor

Random forest may be a hottest and powerful supervised machine learning algorithm capable of performing both classification, regression tasks, that operate by constructing a mess of decision trees at training time and outputting the category that's the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction. Random decision forests correct for decision trees habit of over fitting to their training set [3]. In this study, the data sets considered are rainfall, perception, production, temperature to construct random forest, a collection of decision trees by considering two-third of the records in the datasets. These decision trees are applied on the remaining records for accurate classification. The resultant training sets are often applied on the test data for proper prediction of crop yield supported the input attributes [2]. RF algorithm was used to study the performance of this approach on the dataset. decision trees by considering two-third of the records in the datasets. These decision trees are applied on the remaining records for accurate classification. The resultant training sets can be applied on the test data for correct prediction of crop yield based on the input attributes [1]. RF

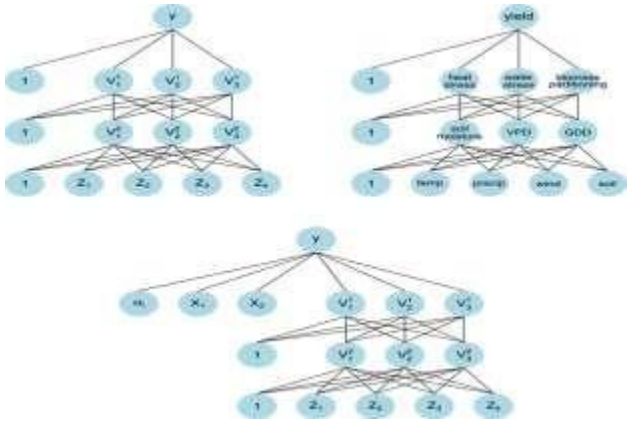

Fig 1: Random Forest Regressor Model

Random forests square measure associate ensemble learning methodology for classification, regression and different tasks, that operate by building a mess of call trees at coaching time and outputting the category that is the mode of the categories or mean prediction of the individual trees. Random call forests correct for call tree custom of over fitting to their coaching set. the primary rule for random call forests was created by Tin Kam Ho victimization the random mathematical space methodology, which, in Ho's formulation, could be a thanks to implement the "stochastic discrimination" approach to classification projected by Eugene Kleinberg. associate addition of the rule was developed by Leo Breiman and Adele bargainer, and "Random Forests" is their trademark. The addition combines Breiman's "bagging" plan and random choice of options, introduced 1st by Ho and later severally by Amit and Geman so as to construct a group of call trees with controlled variance. though random forests are naturally designed to figure solely with third-dimensional information, it's been shown that one may use them for random objects with the employment of solely pairwise similarities between objects.

Random Forest might be a trademark term for associate ensemble of call trees. Random Forest is assortment of call trees (so called Forest). To classify a replacement object supported attributes, every tree offers a categoryification which we are saying the tree votes for that class.
The yield of those crops was tabulated for continuous 20 years by collecting the knowledge from secondary sources. Similarly for the corresponding years climatic parameters such as Rainfall, Maximum & Minimum temperature, Potential Evapo transpiration, Cloud cover, Wet day frequency were also collected from the secondary sources. The methodology adopted for analysis includes for values above the threshold were considered as one child and the remaining as another child. It also handles missing attribute values. In pseudo code, the overall algorithm for building decision trees is:

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

- Check for base cases
- for every attribute a : find the normalized information gain from splitting on a
- Let a_best be the attribute with the very best normalized information gain

• Create a choice node that splits on a_best
• Recurse on the sublists obtained by splitting on a_best, and add those nodes as children of node. Let S be a set of training samples, where the class label of each sample is known. Each sample is in fact a tuple. One attribute is used to determine the class of training samples. Suppose that there are m classes. Let S contain si samples of sophistication Ci for I= 1.., m. An arbitrary samples belongs to class Ci with probability si/s, where s is the total number of samples in set S. The expected information needed to classify a given sample is

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

The information gain obtained by this portioning on A is defined by.
The accuracy obtained is 92.87%

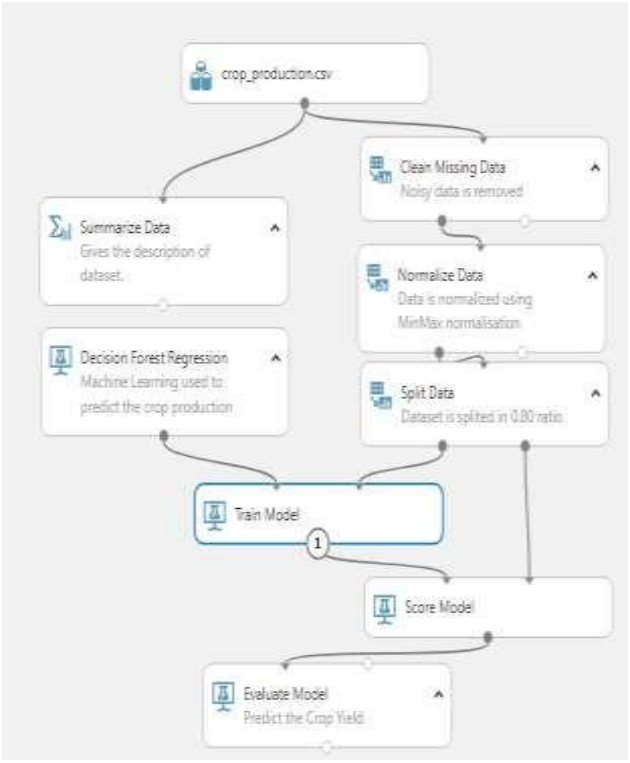

Fig 2: Crop Yield Prediction Flowchart

3. Results

From the analysis of those articles, it had been found that eight ML models are implemented in total. More specifically, five ML models were implemented within the approaches on crop management, where the foremost popular models were ANNs (with most frequent crop at hand—wheat). In livestock management category, four ML models were implemented, with hottest models being SVMs (most frequent livestock type at hand—cattle). For water management especially evapotranspiration estimation, two ML models were implemented and therefore the most often implemented were ANNs.

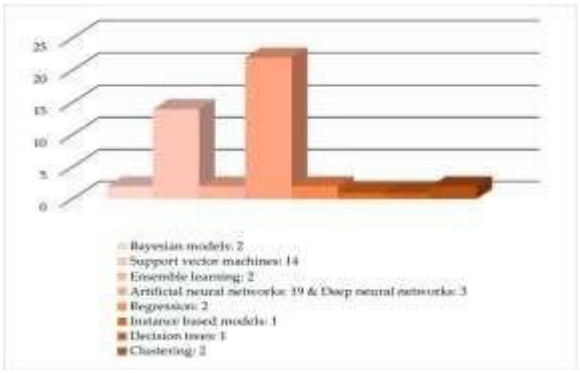

Fig 3 : Comparison of different machine learning algorithms

4. Conclusion and Future Work

The proposed system takes into consideration the data related to soil, weather and past year production and suggests which are the best profitable crops which can be cultivated in the a propose environmental condition. As the system lists out all possible crops, it helps the farmer in deciding of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.
In the future, all farming devices are often connected over the web using IOT. The sensors can be employed in farm which will collect the information about the current farm conditions and devices can increase the moisture, acidity, etc. accordingly. The vehicles used in farm like tractor will be connected to internet in future which will, in real time pass data to farmer about crop harvesting and the disease crops may be suffering from thus helping the farmer in taking appropriate action. Further the simplest profitable crop also can be found in light of the monetary and inflation ratio.

5. Screenshots

Fig 4: Weather Forecast

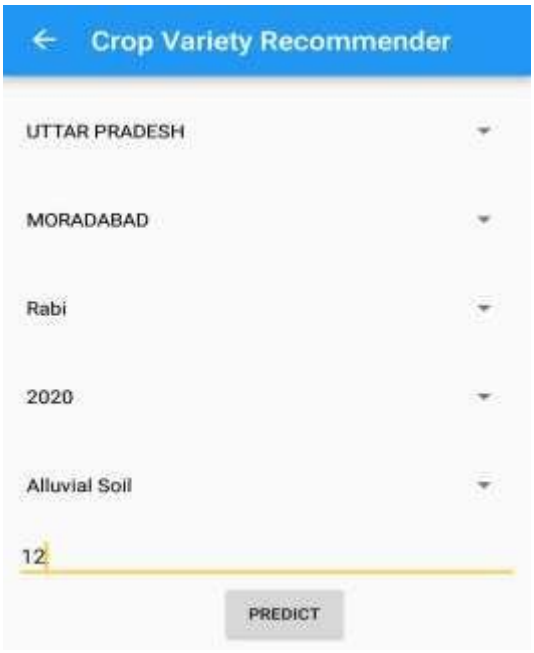

Fig 5 : Crop Variety Recommender Module

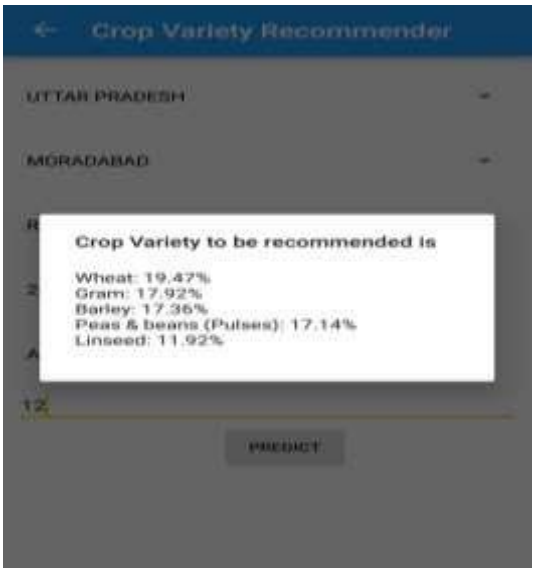

Fig 6 : Crop Variety Recommebded to be are shown

References

[1] Plamondon, Réjean, and Sargur N. Srihari. "Crop Yield Prediction System: a comprehensive survey." Pattern Analysis and Machine Intelligence, IEEETransactions on 22.1 (2000): 63-84.

[2] Madhvanath, Sriganesh, and Venu Govindaraju. "The role of holistic paradigms in precision Agriculture." PatternAnalysis and Machine Intelligence, IEEE Transactions on 23.2 (2001): 149-164.

[3] Park, Jaehwa, Venu Govindaraju, and Sargur N. Srihari. "Crop Prediction in a hierarchical feature space." *Pattern Analysis andMachine Intelligence, IEEE Transactions on* 22.4 (2000):400-407.

[4] Cohen, G., Afshar, S., Tapson, J., & van Schaik, "A Weather and climatic changes affecting modern farming", 2017.