Smart Farming using Machine Learning and Data Analytics

Shardul Pathak[1], Sagar Majgude[1], Sagar Maske[1], Aneesh Sakure[1], Nirmit Singhal[1], Yashwant dongre[2]

[1]*B. Tech Computer Science Engineering, Vishwakarma Institute of Information Technology, Maharashtra, India*

[2]*Assistant professor, Computer Engineering, VIIT, Pune, Maharashtra, India*

Abstract *- 'Smart farming using machine learning and data analytics' is a POC intended for betterment of farmers and agriculture in India. Mentioned in the report are the exact goals to be fulfilled by this project – namely, recommendation of the top n crops based on soil properties and environment factors. These take into consideration all the nutrients and micronutrients (around 12 of them) with reference to the soil health card. From these top n crops recommended, the farmer will also be able to know the economically beneficial crop where the model implements price prediction module for the crops. Also a year round plan will be provided to farmer to maximize his cultivation and earnings through CSM. Besides generic information regarding government schemes or soil testing labs will be provided. The researcher along with these functionalities can access certain datasets that he wants to explore. The researcher can also study the impact of various algorithms on varies data. An add on 'what if' scenario will help not only the farmer but also the researcher to play around with various factors and see the variations in results and proceed with decision making accordingly.*

Keywords: Micronutrients, Farmer, Algorithms.

1. Introduction

'Smart farming using machine learning and data analytics' is a paper intended studying for betterment of farmers and agriculture in India. Mentioned in the report are the exact goals to be fulfilled by this project – namely, recommendation of the top n crops based on soil properties and environment factors. These take into consideration all the nutrients and micronutrients (around 12 of them) with reference to the soil health card. From these top n crops recommended, the farmer will also be able to know the economically beneficial crop where the model implements price prediction module for the crops. Also a year round plan will be provided to farmer to maximize his cultivation and earnings through CSM. Besides generic information regarding government schemes or soil testing labs will be provided. The researcher along with these functionalities can access certain datasets that he wants to explore. The researcher can also study the impact of various algorithms on varies data. An add on 'what if' scenario will help not only the farmer but also the researcher to play around with various factors and see the variations in results and proceed with decision making accordingly.

1.1. Methodologies of Problem Solving

Recognizing the need: Due to the unpredictable environmental factors or price patterns, farmers are unable to decide which crop to cultivate so as to maximize profit. Even if the crop suits his land a lot of it goes waste if there is no market demand. Thus a model was needed to guide the farmer considering his land and environment and price factors. Collection of information and data: Data required in the project was to train the module for prediction. The data is created by the project members from knowledge set – Various textbooks, Agriculture College Pune's App, Soil Health cards study, etc. Price data was taken from government sites. Creating other alternative solutions: From user's point of view, he has been given the flexibility to choose amongst the various solutions provided by app – single or alternative crop, or a year round plan. From developing point more than 4-5 algorithms have been explored, considered, implemented to provide alternative solutions. Evaluating the consequence of different solutions: User might choose a solution as per his necessity. Algorithms were evaluated for accuracy and mean errors to improve accuracy and reduce errors. Deciding and specifying final best solution: Farmer should go for economically beneficial or year round plan. Algorithm with highest accuracy used for base model. Researcher can refer to data he needs.

2. Background and Domain

India being an agro-based country, its economy mainly depends on agriculture produce growth and related agro-industry products. Indian agriculture largely depends on the rainfall which is highly unpredictable. Farming yield is also largely dependent on diverse and multiple soil parameters, like Nitrogen, Phosphorus, Potassium, Crop cycle, Soil moisture, soil pH, temperature and climatic aspects like temperature, rainfall, etc. Nowadays India is very rapidly progressing towards technical innovation and development. Thus, technology will be beneficial as well as blessing to agriculture which will enhance crop productivity which will result in better yields to the farmer.

2.1 Agriculture

Different types of soils have various mineral contents and every crop require

multiple mineral components for its better growth. Each and every soil has some specific characteristics and is suitable to grow only a particular number of crops. Therefore an agronomist should be known about the type of soil he has so that he could cultivate better crops. Another important issue that is famers have to face is the uncertainty in the market demand and prices from the time of sow to the time of harvest.

2.2 Machine Learning

Machine Learning is a section of Computer Science where new developments evolve at recent times, and also helps in automating the evaluation and processing done by the mankind, thus reducing the burden on the manual human power. It is a type of Artificial Intelligence that provides devices with the ability to learn without being separately programmed. It gives focus on the development of computer programs that can change when exposed to new data. Finding out the suitable crops based on the soil's appearance becomes tedious for novice farmers. There also exists a need to prevent the agricultural decay.

2.3 Data Analytics

It is a method of analyzing, cleaning, transforming & creating models of data with the aim of finding meaningful and useful information, concluding that information and supporting decision making. Data analytics projects and initiatives support businesses increase revenues, help in improving operational and functional efficiency, responds quickly to emerging market trends and gain a competitive edge over rivals, optimizes marketing campaigns and customer service efforts.

3. Software Requirements

3.1 Analysis Models: SDLC

Software development life cycle for the proposed project went through following phases: Requirements Gathering: The requirements for the project started right from the basics to see if the problem is really worth solving. Thus various agriculture colleges were visited to gather the requirements. The requirements mostly included precision in the prediction as, though existing models can be seen, most of them fail to provide accurate results. The data required for the same was gathered in the form of knowledge set through agriculture college textbooks and official app. The technical requirements of platform, technologies, etc. were decided through discussions with the guide. Analysis: The results from the existing models failed due to the uncertainty in environmental conditions and lack of parameters considered for prediction. Thus it was decided to consider parameters that would yield a better result. Dataset was generated from the knowledge set by study and analysis. A lot of papers through literature survey also helped to analyses the gaps that could be worked upon for a better model. Plan and Design: Thus the software design was generated that included the software architecture, all the necessary UML diagrams, and schemas for data and the modules that would be provided through the web app. Key deliverables of the app were decided and usage of algorithms, technologies, platforms, language, database, etc. were also decided. The layout of the app and its interface were proposed. A plan was thus sketched. Build: Actual coding required was done. This included machine learning algorithms execution, front end codes – for static and dynamic web pages, machine learning codes API generation, etc. between devices. It also enables better utilization of the grid by detecting the device's efficiency at a particular time. Remote sensing is one of the

important wireless measures of Smart Grid Applications. Test: The generated web app was tested and validated for various situational inputs. This included mostly usability testing, security testing, maintainability testing, scalability testing, login testing, database testing, compatibility testing, etc. Deploy: Presently the app is deployed as client server architecture with clients residing in the same network as that of the server. Further deployment on cloud for access over the internet is planned. Maintain: Check for bugs and tackling them is planned when will be needed. The system would be maintained for its availability, security, usability, scalability, and any updates based on users demand would be tried to fulfil.

4. System Design

4.1 System Architecture Diagram

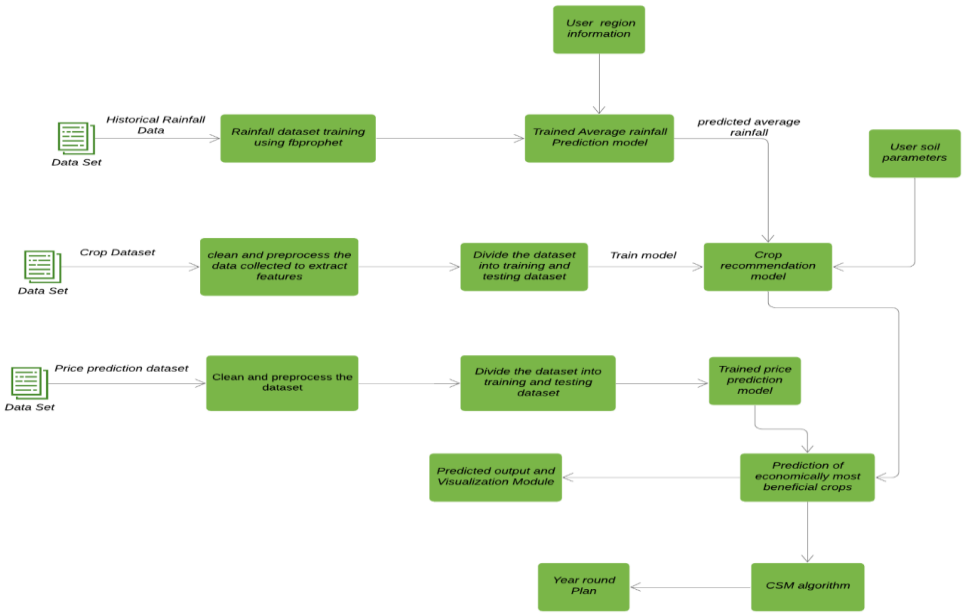

Figure 4.1 System Architecture

4.2 System Data Flow Diagram:

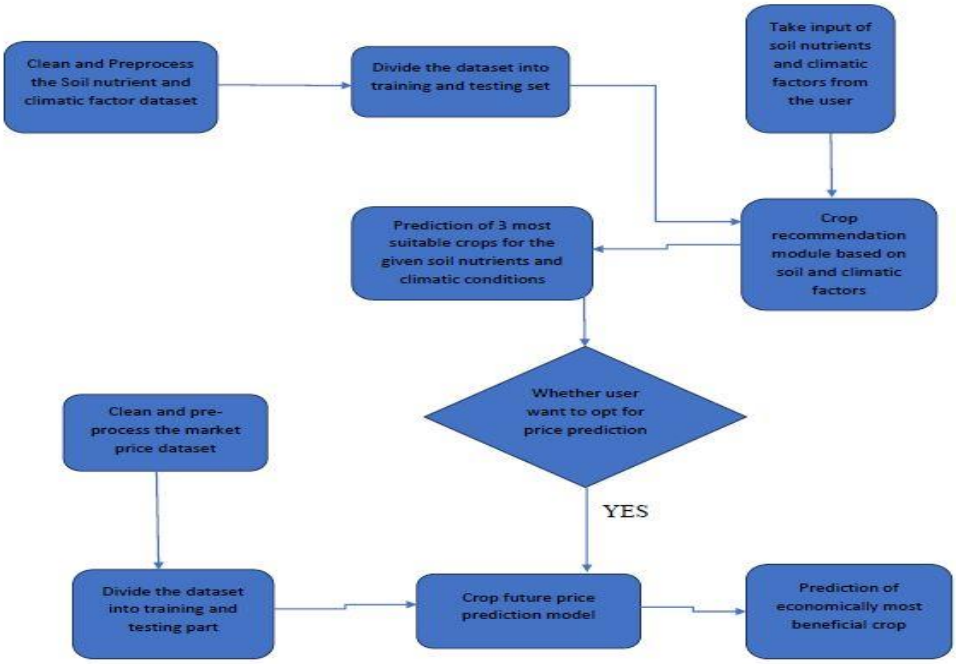

Figure 4.2 System Data Flow Diagram

5. Datasets

5.1 Soil nutrient and climatic factors for crop dataset:

Many researchers nowadays are working on crop recommendation and crop prediction problem despite of that finding data related to this was the biggest hurdle we faced during our project. The major focus of our project was to include climatic conditions as well as all the major soil parameters for crop recommendation because no model yet includes them all together for multiple crops. Hence, no readymade dataset was available to us. We made use of the available knowledge sets to build our final dataset which includes all the necessary soil parameters as well as climatic factors.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	crop_name	Soil_Type	Duration	Temp	Rainfall (cm)	pH	EC (dS/m)	Organic_Carbon (%)	N (Kg/ha)	P_2O_5 (Kg/ha)	K2O (Kg/ha)	S (ppm)	Zn (ppm)	B (ppm)	Fe (ppm)	Mn (ppm)	Cu (ppm)
	Rice	clay	130	26	6	5.5		H	H	H	H	S	S	S	S	S	S
	Rice	clay	130	26	6	5.6		H	H	H	H	S	S	S	S	S	D
	Rice	clay_loam	130	26	6	5.7		H	H	H	H	S	S	S	S	S	S
	Rice	clay	130	26	6	5.8		H	H	H	H	S	S	S	S	S	D
	Rice	clay_loam	130	26	6	5.9		H	H	H	H	S	S	S	S	S	D
	Rice	clay	130	26	6	6		H	H	H	H	S	S	S	S	S	S
	Rice	clay_loam	130	26	6	6.1		H	H	H	H	S	S	S	S	S	D
	Rice	clay	130	26	6	6.2		H	H	H	H	S	S	S	S	S	S
0	Rice	clay_loam	130	26	6	6.3		H	H	H	H	S	S	S	S	S	D
1	Rice	clay	130	26	6	6.4		H	H	H	H	S	S	S	S	S	S
2	Rice	clay	130	26	6	6.5		H	H	H	H	S	S	S	S	S	D
3	Rice	clay_loam	130	27	6	5.5		H	H	H	H	S	S	S	S	S	S
4	Rice	clay_loam	130	27	6	5.6		H	H	H	H	S	S	S	S	S	S

Figure 5.1 Crop Dataset

The next important question to discuss is why we have chosen only these particular soil parameters. During our research work in the early phase of our project we came across the concept of soil testing. We came to know that lot of soil testing laboratories have been set up by government as well as by some private institutions which performs the analysis of soil and give details of that analysis in the form of **SOIL HEALTH CARD** to famers.

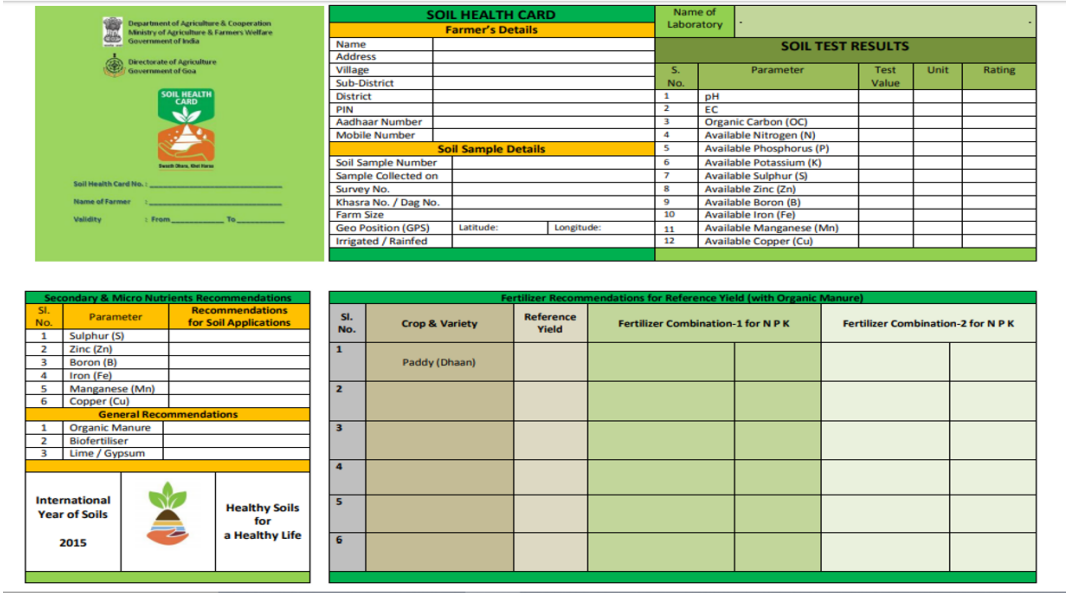

Figure 5.2 Soil Health Card

5.2 Soil Health Card:

It is Indian Government's scheme promoted by the Department of Agriculture and Co-operation under the Ministry of Agriculture and Farmer's welfare. This scheme is being implemented by the Department of Agriculture of all the State and Union Territory Governments. A Soil Health Card gives farmer soil nutrient status of his holding and advices him on the usage of fertilizers and also the needed soil amendments that he should apply to maintain soil health. A printed report of Soil Health Card has the status of soil having 12 parameters, namely N, P, K (Macro-nutrients) Fe, Cu, S, Zn, Mn, Bo (Micro-nutrients); and EC, pH, OC (Physical parameters). We wanted to include all the major factors that influence the crop suitable for particular soil type hence we had to consider all these 12 factors along with type of soil average rainfall, temperature of that region.

5.3 Market Price Prediction Dataset:

For future price prediction, the dataset is collected from agmarknet.nic.in which a government website is providing users with prices of crops. It has district wise price distribution having minimum price, maximum price and modal price.

Sl no.	District Name	Market Name	Commodity	Variety	Grade	Min Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal Price (Rs./Quintal)	Price Date
1	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1711	2500	2105	31-Dec-19
2	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	2031	2342	2186	30-Dec-19
3	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1697	2300	1998	28-Dec-19
4	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1726	2122	1924	27-Dec-19
5	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1700	2000	1850	26-Dec-19
6	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1775	2075	1925	24-Dec-19
7	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1800	2053	1926	23-Dec-19
8	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1850	2500	2175	21-Dec-19
9	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1675	2222	1948	20-Dec-19
10	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1661	2400	2030	19-Dec-19
11	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1650	2400	2025	18-Dec-19
12	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1661	1863	1762	17-Dec-19
13	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1661	2200	1930	16-Dec-19
14	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	2050	2300	2175	14-Dec-19
15	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1651	2200	1925	13-Dec-19
16	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1600	1625	1612	12-Dec-19
17	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1900	2400	2150	11-Dec-19
18	Ahmednagar	Ahmednagar	Bajra(Pearl Millet/Cumbu)	Other	FAQ	1800	2475	2137	10-Dec-19

Figure 5.3 Price Dataset

5.4 CSM Dataset:

The third dataset that we are using is csm_dataset for Crop Selection Method algorithm. This dataset contains the sowing period, harvesting period, growth period, and predicted yield rate of crop.

	A	B	C	D	E	F
1	crop_name	Sowing_period	Harvesting_period	Growth_day	Predicted_yield_rate	
2	Rice	June	Sept	4	2000	
3	Rice	July	Oct	4	2000	
4	Soybean	June	Oct	5	1264	
5	Soybean	July	Nov	5	1264	
6	Sweet_potato	July	Oct	4	800	
7	Sweet_potato	Aug	Nov	4	800	
8	Arhar	July	Dec	6	1359	
9	Arhar	Aug	Jan	6	1359	
10	Castor_seed	July	Feb	8	1064	
11	Castor_seed	Aug	Mar	8	1064	
12	Wheat	Oct	Feb	5	788	
13	Wheat	Nov	Mar	5	788	
14	Potato	Oct	Dec	3	1650	
15	Potato	Dec	Feb	3	1650	
16	Ratoi	Oct	Dec	3	1472	
17	Ratoi	Nov	Jan	3	1472	
18	Sarso	Oct	Dec	3	1800	
19	Sarso	Nov	Jan	3	1800	
20	Lineseed	Oct	Feb	5	1182	
21	Lineseed	Nov	Mar	5	1182	

Figure 5.4 CSM Dataset

6. Algorithms:

6.1 Machine Learning

Machine Learning is a section of Computer Science where new developments evolve at recent times, and also helps in automation of the evaluation and processing done by the mankind and reduces the burden on the manual human power. It is a type of Artificial Intelligence that provides devices with the ability to learn without being separately programmed. It gives focus on the development of computer programs that can change when exposed to new data. Finding the suitable crops based on the soil's appearance becomes tedious for novice farmers. There also exists a need to prevent the agricultural decay.

6.2 Neural Networks

The Neural Networks Model (NNM) is used to consider the functionality of non-linear data. The model shows the capability to learn as it follows comparable data processing structure. These techniques provide successful results by applying on many problems like classification, control and prediction like a biological brain. The model is dissimilar to both classification model as well as decision tree due to its factor of likely hood prediction. The neural network has many techniques which has merits and demerits. The research suggests neural network to be better than decision tree and regression analysis model used in churn prediction.

6.3 Decision Tree

The decision tree is one of the most prominent predictive models that are used for the churn prediction based on classification. The decision tree has of two steps: 1. Tree building - The training data set is recursively partitioned according to values of the attributes. This process continues till there are no partitions left having identical values. 2. Tree pruning - Here some values may be removed from the data which is noisy data. The largest found out error rate branches are selected and further removed in pruning. To predict accuracy and reduce complexity of decision tree is thus called tree pruning.

6.4 Data Analytics

It is a method of analyzing, cleaning, transforming & creating models of data with the aim of finding meaningful and useful information, concluding that information and supporting decision making.Data analytics projects and initiatives support businesses increase revenues, help in improving operational and functional efficiency, responds quickly to emerging market trends and gain a competitive edge over rivals, optimizes marketing campaigns and customer service efforts.

7. Conclusion

Summing up, as this paper is an attempt is made to bridge the gap between labor, finance, land etc. invested in agriculture and yet its lower contribution to GDP. The datasets needed for the project have been gathered. It includes datasets with fields of land properties, environmental factors as well as the future price consideration of crops. Interactions with various agricultural colleges, labs and experts have been done for the same. As mentioned, data analytics and machine learning are the main domain areas focused on technically.

8. References

[1] Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique [IEEE 2015] - Rakesh Kumar1, M.P. Singh2, Prabhat Kumar3 and J.P. Singh.

[2]Prediction of Future Market Price for Agricultural Commodities [International Journal of System and Software Engineering 2015] - Sagar Pathane, Uttam Patil, Nandini Sidnal.

[3] CROP PREDICTION USING PREDICTIVE ANALYTICS [IEEE 2017] - P. S. Vijayabaskar, Sreemathi.R, Keertanaa.E.

[4] Predictive Analysis to Improve Crop Yield using a Neural Network Model [IEEE 2018] - Shruti Kulkarni, Shah Nawaz Mandal, G Srivatsa Sharma, Monica R Mundada, Meeradevi.

[5] Recommendation System for Crop Identification and Pest Control Technique in Agriculture [IEEE 2019] - Avinash Kumar, Sobhangi Sarkar and Chittaranjan Pradhan.

[6] Crop Yield Prediction and Efficient use of Fertilizers [IEEE 2019] - S.Bhanumathi, M.Vineeth and N.Rohit.

[7] Decision Making Support System for Prediction of Prices in Agricultural Commodity [IEEE 2019] - Aman Vohra1, Nitin Pandey2, S.K. Khatri3.