ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse501032021.pdf
https://doi.org/10.30534/ijatcse/2021/501032021

Agrobot in field of Machine Learning

Putul Siddharth[1], B V A N S S Prabhakar Rao[2]
[1]Department of Computer Science and Engineering, Vellore Institute of technology, Chennai, India,
sidputul1997@gmail.com
[2]prabhakarrao@vit.ac.in

ABSTRACT

Farming is still one of the largest sectors of India from a financial standpoint as well as employing almost two-thirds of our working population. Yet the farming industry suffers from not being attuned with modern technological advancements. Hence farming industry has not reached its optimum potential and has a lot of areas for improvement. Our project tries to tackle some of these problems. The project's main focus is on a crop recommendation system that was built using Machine Learning techniques. Along with this, we have worked to make a website as a UI for our recommendation system. Finally, we have also added a lot of agriculture-related infotainment as there is a lot of misinformation related to agriculture and also to pique the interest of the user in the field of modern agriculture. Through the whole process of implementing this project, our team learned about Machine Learning, Web Development, Content Creation, and most importantly the scientific method of Agriculture for the 21st century. Our design ideology revolved around making our system a holistic place for anyone who wants to get an introduction to modern farming. The germination of the project is from the need to modernize agriculture and thus the project intends to tackle issues from the grass root.

Key words: Decision tree, KNN, Logistic regression, Naïve Bayes, Random forest, Recommendation system, Support vector machine.

1. INTRODUCTION

Agriculture is one of India's most important industries, but more can be done to carry it into the twenty-first century by making it more scientific and industrially precise. There are many approaches to achieving this aim, and our project addresses some of them. Our framework aims to assist in crop recommendation, as well as provide agriculture-related information.

India, as a predominantly agricultural nation, requires agricultural innovation. One way we're attempting to achieve this is by offering a crop recommendation system to a farmer or other consumer, allowing them to choose the best crop to grow based on their area and type of soil. This is intended to eliminate inefficiency in the agriculture sector, which is already an unorganized sector.

From the application, the project will assist farmers in making better and faster decisions about which crop to grow and in risk management. Other features, such as blog entries, crop information, agricultural loan information, and vacant land information would also be able to pique the attention of both novices and experts.

1.1 Objectives

Following are the objectives of the paper:

- To collect agricultural datasets that has soil chemistry and location-related information for different crops. Doing data processing to help in building the recommender systems. Carrying out dataset analysis on various ML models.
- Build crop predictor system using soil chemistry dataset with the help of Random Forest as our ML algorithms.
- Build crop recommender system using location-based dataset and crop-based dataset with the help of KNN algorithm.
- To host a website to make the system user-friendly.
- Implementing account creation system in website.
- Providing agriculture-related blog posts and other information for users.

2. LITERATURE SURVEY

The paper[1] focused on the need for crop yield prediction in the modern-day world and uses various lazy and eager learner algorithm to tackle this issue. They aim at finding a suitable data model that gives the highest accuracy against their dataset.

The paper[2] focused on crop recommendation systems using data mining techniques. A system is developed in which the user has to enter the soil nutritional attributes through an interface and the system will map it to the crop dataset with the data values collected over the years and thus a corresponding class of crop is predicted using the Naive Bayes Algorithm. There will be an admin panel that is responsible for handling and managing the entire application. The system has been developed to provide benefits to the farmer in terms of crop and risk management.

The paper[3] emphasizes the need for a Fuzzy Model-based crop recommender framework for farmers. Here the dataset is prepared through Questionnaire Assessment from the district Jaipur of state Odisha. The questionnaire consists of twenty

Questions on different attributes of soil and weather parameters along with rice production. This questionnaire was asked from the maximum qualified people of the district in their native language. The system asks for location, crop preferences, and season as input. Now the system mapped this set of input to the dataset to check the yield of the corresponding crop. The system used three input two output Mamdani Fuzzy inference models for prediction. The study was limited only to rice production and has not taken many weather attributes into account.

The paper[4] is about a crop monitoring framework that uses machine learning and the Internet of Things. The soil parameters are collected with the help of sensors then these collected data are uploaded into NodeMcu which has a built-in Wi-Fi module. Now, these uploaded data are sent to the ThingSpeak cloud for generating responses to the android application. Finally, the responses generated from the cloud are sent to the ML model for triggering appropriate crop recommendations. The system also has a price prediction section that estimates the price of the yield using the linear regression model.

3. METHODOLOGY

3.1 Dataset Collection

For purpose of this paper, we searched for datasets that would allow us to make a crop recommender with multiple attributes. We ideally wanted a single dataset that would have soil chemistry data and location data respective to different crops. After searching the web extensively for datasets matching our requirements and after analyzing different datasets, we have settled on three different datasets which will be discussed below.

3.1.1Soil chemistry based dataset

The soil chemistry dataset, a part of which is displayed in Table 1. is the main dataset for the project. The dataset has 2200 rows and 8 columns.

Table 1: Soil chemistry based dataset

N	P	K	Temperature	Humidity	pH	Rainfall	Crop
90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	rice
71	54	16	22.61359953	63.69070564	5.749914421	87.75953857	maize
40	72	77	17.02498456	16.98861173	7.485996067	88.55123143	chickpea
13	60	25	17.13692774	20.59541693	5.68597166	128.256862	kidneybeans
27	72	17	28.98039357	57.23265151	6.347929353	120.7435664	pigeonpeas
3	49	18	27.91095209	64.70930606	3.692863601	32.67891866	mothbeans
36	55	20	27.01470397	84.34262707	6.635968698	55.296354	Mungbean
50	55	16	28.81460716	65.33538112	7.581442888	62.26242533	Blackgram
11	74	17	21.36383757	69.92375891	6.633864582	46.6352865	Lentil
15	11	38	23.12808226	92.68328358	6.630646083	109.3930157	Pomegranate
90	92	55	27.00932084	80.18546798	6.13465588	97.32531705	Banana
18	19	27	27.75518664	52.34605806	4.772385986	94.11213345	Mango
32	141	204	8.825674745	82.89753705	5.536645599	67.235765	Grapes
105	30	50	25.29954705	81.77527562	6.37620108	57.04147057	Watermelon
82	18	48	29.09588297	94.16748386	6.159050816	26.70581328	Muskmelon
22	144	196	21.91191314	91.68748063	6.499226821	117.0761277	Apple
20	7	9	29.47741671	91.57802915	7.129136941	111.1727497	Orange
56	57	48	31.56213762	93.0484859	`6.506120752	63.62250788	Papaya

It has the following attributes:

- Nitrogen (N): Part of the "Enormous 3(NPK)" supplements needed by plants for their food. Any plant ingests Nitrogen the most of all nutrients. Nitrogen is fundamental in the development of proteins which are urgent for the development of any plant. Soil Nitrogen esteem is accordingly fundamental in deciding the kind of harvests that could develop there as various yields require a shifting level of nitrogen.
- Phosphorus (P): Second of the “Big 3(NPK)” nutrients required by plants for their sustenance. Phosphorus is used by plants to store energy and also to do photosynthesis. It also aids in plant growth and development so it becomes crucial for any crop. As mentioned for N, soil P values also differ from crop to crop.
- Potassium (K): Third of the “Big 3(NPK)” nutrients required by plants for their sustenance. Potassium is responsible for plant immunity and increasing yield. K helps strengthen root systems in dry/cold conditions. As stated for N & P, every crop needs a different amount of Potassium to thrive.
- pH: pH can be considered an important factor as it affects many other soil attributes and processes that influence plant growth. The behavior of microorganisms, as well as the solubility and availability of nutrients, are all pH-dependent processes so when the pH of the soil is about seven, most crops thrive.
- Temperature: Temperature is perhaps the main variables in cultivating as irregular characteristics in it can prompt demise of harvests, failure of manures among different issues. Right temperature is additionally crucial for give better yield quantitatively and subjectively.
- Humidity: A feature that is indirectly dependent on air temperature thus making it very important for crop production. More humidity tends to cause foliar or root diseases in plants thus increasing pesticide costs and decreasing yield. On the other hand, low humidity leads to slow growth and more leaf drops.
- Rainfall: Another very important environmental feature for crop production. Less of it can starve crops or increase production costs as artificial watering would have to be ramped up. More of it would mean the inundation of crops or dilution of pests and fertilizers making crops vulnerable to diseases.
- Crop: The dataset maps specific crops like rice with the above 7 attributes. Since most crops have a tolerance range in each of the 7 attributes that is why there are multiple entries per crop.

This dataset was used to build the prediction system. This dataset has 22 varieties of crop like rice, watermelon, etc. with 100 entries for each type to get a better model for each crop and thus making the dataset more uniform avoiding any bias due to less data for any specific crop.

3.1.2 Crop and Location based dataset

Table 2: Crop based dataset

Crop	N	P	K	pH
Rice	80	40	40	5.5
French Beans(Farasbi)	90	125	60	5
Green Peas	40	35	55	6
Jambun(Syzygium cumini)	10	10	10	4.2
Tamarind	20	15	25	3.1
Banana	100	75	50	6.5
Ragi	50	40	20	5.5
Lima beans(Pavta)	40	60	20	5
Soyabean	20	60	20	5.5
Horse Gram(kulthi)	20	60	20	6
Coriander seeds	90	20	20	6.5
Garcinia indica(kokam)	40	25	50	4.2
Lemon	20	10	10	3.1
Figs	20	60	20	5
Grapes	20	125	200	4

The crop-based dataset and location-based dataset are the other dataset used for the project.The initial aim was to find one dataset that also had location-related information along with soil chemistry and environmental factors related to crops.

Table 3: Location based dataset

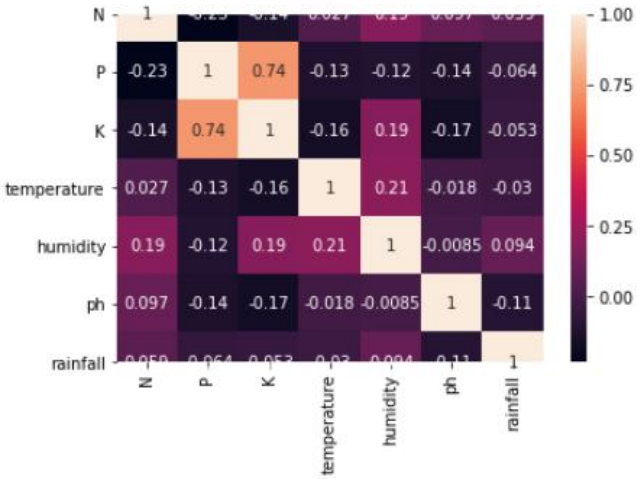
Location	Y	X	N	P	K	pH
Khudapura	14.4212044	76.5545525	104	23	106	6.98
Manamainahatti	14.4251505	76.5226052	125	28	125	6.77
Turuvanur	14.4003794	76.4305865	103	13	159	5.98
Ganjigatte	14.2341279	76.1157275	63	9	159	8.14
Gyarahalli	14.2285746	76.1075307	54	10	159	8.16
Muthugaduru	14.2197345	76.1154271	103	20	170	7.54
Saasalu	14.1993182	76.1131633	103	15	169	6.43
Daginakatte	14.2005352	75.781889	103	17	166	7.26
Hunesehalli	14.2148461	75.7170703	74	10	179	8.14
Basavapura	14.2251706	75.7335002	85	21	184	7.27
Holemadapura	14.2192406	75.6751802	79	10	154	7.62
Halivana	14.3233453	75.7592645	95	9	198	5.26
Chikkathammanahalli	14.3079906	75.7423603	113	14	190	6.29
Dibbadahalli	14.3363521	75.7415808	76	9	151	5.83
Anjaneya Swamy Temple	14.4354257	75.7647429	51	9	169	8.3
Haralahalli	14.3326983	75.7777101	55	9	150	8.3

pH, Nitrogen (N), Phosphorus (P), and Potassium (K) are the seed parameters that are used to classify the crop. Crop parameters are used to characterize a diverse range of crops grown in India's fertile lands. The soil dataset includes data samples collected from various locations in Chennai and other Tamil Nadu districts. pH, Nitrogen (N), Phosphorus (P), and Potassium (K) is the soil parameters that are used to classify the soil. The data samples were collected from the Tamilnadu-Kancheepuram district's approved horticulture department. To collect feedback from users, the device employs an interface. Depending on the module they choose, the input form can vary.

3.2 Crop Prediction

As we all know that crop prediction is of great use to global food production as many of them rely on accurate prediction to make timely import and export decisions and can make both ends meet. The crop prediction starts with importing the dataset from Kaggle and necessary packages and libraries are installed. Heat map is then used to show the correlation between different attributes of a dataset. Then cleaning the dataset is done in case there are null values present in the dataset.

After that Normalization of the dataset is performed using Standard Scalar to increase the efficiency of our ML model. Finally, we have used different machine learning algorithms like DecisionTreeClassifier, GaussianNB, Support Vector Machine, Logistic Regression, and RandomForestClassifier against our dataset to compare the accuracy and the below table showed the same.



.**Figure 1:** Snapshot of heat map

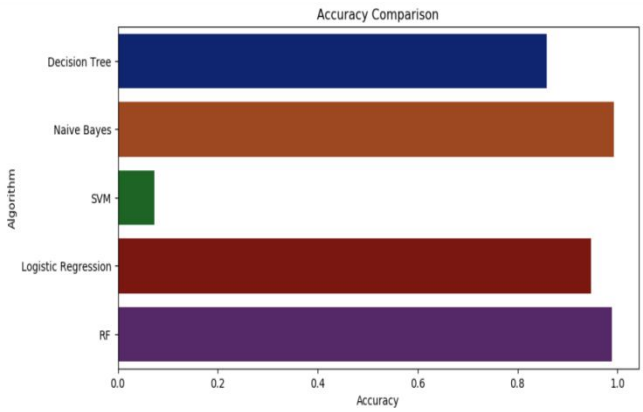


Figure 2: Algorithm vs. Accuracy Comparison

Now we considered Random Forest as our classification algorithm for further prediction and use the pickle module to save our model. So whenever the user enters the required input through an interface, this model is used to predict the crop and display it to the user.

3.3 Crop Recommendation

Along with crop prediction, our project also encourages the recommendation of a set of different crops to the farmers based on two methods either via their city location or via their soil nutrients value. Similar work is done for prediction, which involves dataset import, cleaning, and normalizing the dataset, but for crop recommendations, we used k-nearest neighbors.

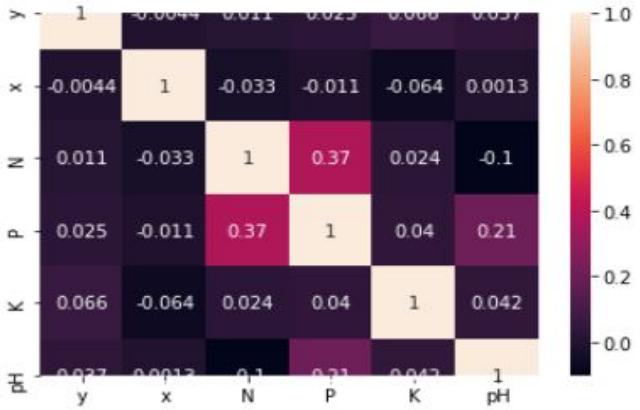


Figure 3: Snapshot of heat map

In the first part, when the user inputs his city location through the interface, the given input will be mapped to the location-based dataset to gather the soil parameters of the corresponding city. Then this list of parameters will be mapped to that crop dataset and with the help of the KNN classification algorithm, this list of the parameter will find themselves a suitable class of crop and thus the result is displayed to the user through an interface.

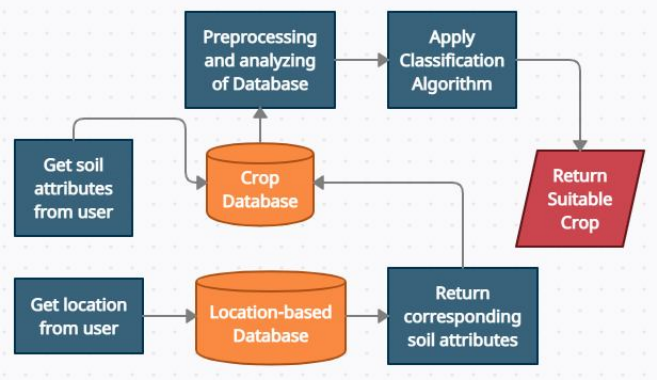


Figure 4: System architecture diagram

In the subsequent part, the user needs to include soil supplements for crop suggestions. The provided input will be saved in a variable, and the crop dataset will be used to map it. With the help of the K-Nearest Neighbor classification algorithm, the input locates itself into a crop class and the same will be displayed to the user using an interface. Subsequently, the user will get a list of crops suggested by the system.

3.4 Agricultural Infotainment

In contrast to other sectors, there is a lack of well-presented agricultural infotainment. As a result, the project aimed to provide blog-style content that any reader could understand and that would allow them to follow or at the very least be aware of modern agricultural methods. In addition, the system provides the consumer with information on crops, agricultural loans, and available vacant land for cultivation.

This was done by creating HTML pages for each page that can be displayed to the user. The pages were stylized using CSS Bootstrap. The contact information entered by the user, blog posts, and vacant land information is stored in PHPMyAdmin Databases.

3.5Architectural Diagram

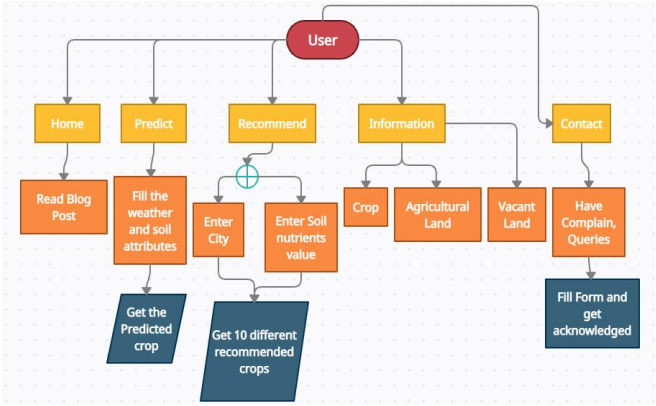


Figure 5: Flowchart of how user uses the system

The user first needs to have an account to access the application. For that, they need to sign up by filling in some of the personal details and set up strong login credentials. Then they can log in by entering correct credentials to utilize different functionalities accessible. At the homepage, the user enjoys the different blog posts related to topics like agro-education, agro-news, etc. Then the user has crop prediction functionality available in which they can get a single predicted crop after entering the required details in the portal. Also, the system provides a crop recommendation facility in which users can get ten different options of the crop suitable for their soil and location.

Now out of the recommended crops if the user is interested in a particular crop but is unaware of its process of cultivation then he can get that knowledge from the information sector of our application. There he can know about health benefits, climate required, soil requirement, land preparation, planting and spacing, irrigation, manures and fertilizers, harvest, yield, and demand of that crop in the market.

Even if a farmer needs some loan to fund his seasonal agricultural operation and other activities but is unaware of the process of getting an agricultural loan, he can get information from the Agricultural Loan sector available in the information menu. There he can know about different types of agricultural loans, agricultural loan interest for different banks in India, the process to apply for an agricultural loan, documents required, and eligibility criteria to borrow an agricultural loan.

The system also proved beneficial for peasants as peasants don't own agricultural land and has to search for that vacant land here and there. So, the system will provide a feature for

the user who wants to lend his land for cultivation on a profit basis. The user needs to fill in some details like owner name, description about land, location, and type of peasants which he needs, contact details, etc. Now the system will broadcast this information to all the users including the peasants. So the one who needs this can contact the owner and help themselves.

Finally, after using the application if the user has some query or complaint, he can get in touch with the admin by just filling in a contact form.

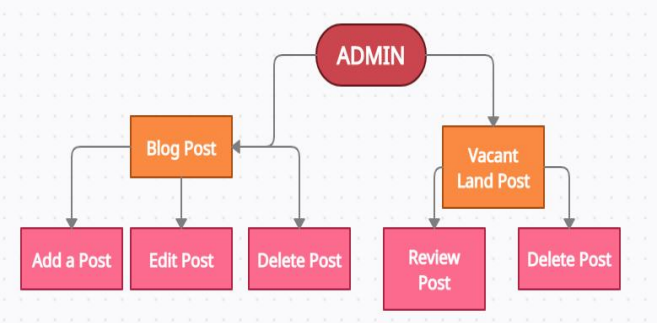


Figure 6: Flowchart of how admin uses the system

There is an admin panel in which the admin has the privilege to add, edit and delete the blog shown on the user homepage. Also, he can review and delete the vacant land post after getting contacted by the corresponding landowner.

For adding or deleting posts our strategy is simple if the serial number is zero then add all form data as a new post in the database but if the serial number is any other number then edit that number of blogs. To add data to the database we will use the "Posts" class in which we pass the object of Sqlalchemy. We will pass form data through arguments and then add the data to the database through "db.session.add ()".

4. RESULTS AND DISCUSSION

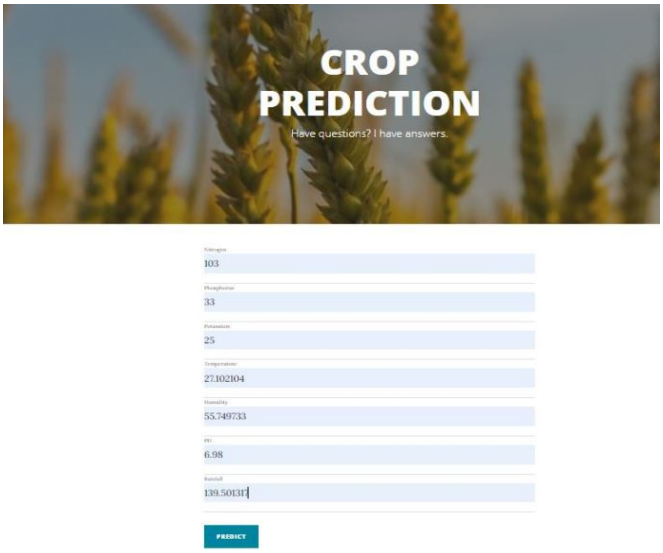


Figure 7: Snapshot of Crop Prediction

Using Random forest as our classification algorithm, a prediction system was built. The datasets used were not as extensive as was expected however they were unbiased towards any specific crop because each of the 22 crops had 100 data instances each. After analyzing different ML algorithms, Random forest was chosen as it gave 98.93 percent accuracy.

For the recommendation system, 10 different suggestions were generated for each entry using the KNN classification algorithm. "predict_proba" function of MultiOutputClassifier was used for this.

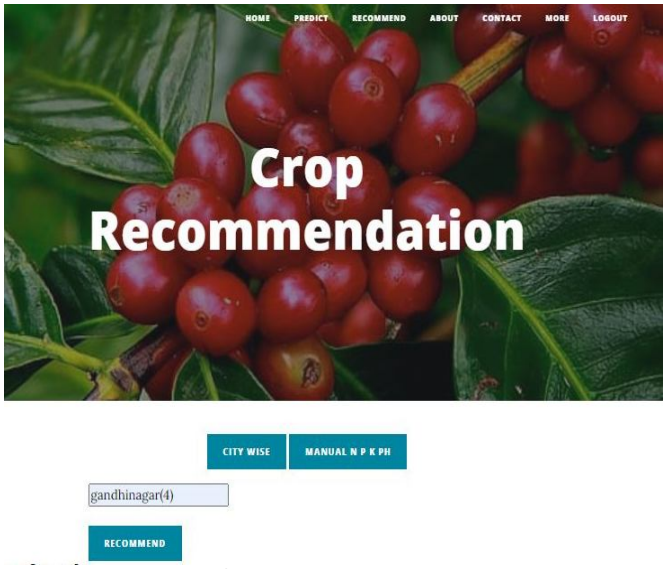


Figure 8: Snapshot of Crop Recommendation widget portal based on location

Figure 9: Snapshot of recommended crops based on location

The most relevant crops are listed above based on the location entered by user in the portal.

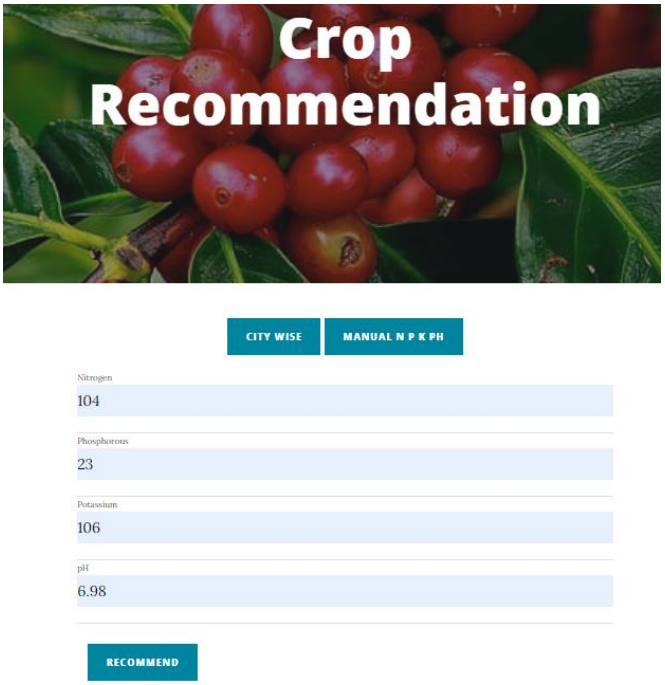


Figure 10: Snapshot of Crop Recommendation widget portal based on soil nutrients value.

Crop found : 10

Crop Name
Bajra(Pearl Millet)
Cluster Beans(Gavar)
Mung beans
Green Peas
Cumin seeds
Guava
Apple
Coconut
Pistachio Nut
Cotton

Figure 11: Snapshot of recommended crops based on soil nutrients value entered by user.

5. CONCLUSION

The recommendation system was created upon a sub-optimal dataset. We are trying to improve the dataset by collecting more data. We are also trying to include parameters so that the dataset mimics the real world much better. Including parameters like crop yield (both by weight and by money), Crops grown before on soil will make the dataset more robust. For now, the ml model works with acceptable accuracy.Due to Covid-19, the lockdown has been imposed all over the country and people are not in a state to move outside their house. So this recommendation system will be able to help farmers to choose the crops to grow with more precision instead of going to any agricultural center to gain information so that they can avoid losses and maximize profits. By the creation of the E-commerce site, I hope casual readers will become interested in agriculture and will thus help in making farming scientific and efficient.

REFERENCES

1. D Ramesh, B Vishnu Vardhan. "**Data Mining Techniques and Applications to Agricultural Yield Data**". *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 2, Issue 9, September 2013, pp. 3477-3480.
2. Bharath Kumar R, Balakrishna K, BencyCelso A, Siddesha M, Sushmitha R, "**Crop Recommendation System for Precision Agriculture**," *International Journal of Computer Sciences and Engineering*, Vol.7, Issue.5, pp.1277-1282, 2019.
3. MadhusreeKuanr, BikramKesariRath, SachiNandanMohanty(2018), ' **Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model**', *International Journal of Engineering & Technology*, 7, (4.15), (2018), 277-280
4. R.Pallavi Reddy, B. Vinitha, K. Rishita, K.Pranavi(2020), ' **Crop Monitoring and Recommendation System using Machine Learning and IOT**', ' *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*', 2278-3075, Volume-9 Issue-9, July 2020.
5. Bhanumathi, M.Vineeth and N.Rohit , "**Crop Yield Prediction and Efficient use of Fertilizers**" , *IEEE International Conference on Communication and Signal Processing*, April 4-6, **(2019)**, pp. 0769-0771.
6. M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour and E. -H. M. Aggoune, "**Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk**," in IEEE Access, vol. 7, pp. 129551-129583, 2019, doi: 10.1109/ACCESS.2019.2932609.
7. J. J. I. Haban, J. C. V. Puno, A. A. Bandala, R. KerwinBillones, E. P. Dadios and E. Sybingco, "**Soil Fertilizer Recommendation System using Fuzzy Logic**," 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020, pp. 1171-1175, doi: 10.1109/TENCON50793.2020.9293780.
8. A.Chandavale, A.Dixit, A.Khedkar and R.B.Kolekar, "**Automated Systems for Smart Agriculture**," *2019 IEEE Pune Section International Conference* (PuneCon), 2019, pp. 1-6, doi: 10.1109/PuneCon46936.2019.9105686.

9. Suchithra M.S and Maya L. Pai, "**Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters**", China Agricultural University. Production and hosting by Elsevier **(2019)**, pp. 1-6.
10. Sadia Afrin, Abu Talha Khan, MahrinMahia, RahbarAhsan, MahbuburRahmanMishal, Wasit Ahmed and Rashedur M. Rahman, "**Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh**", Journal of IEEE **(2018)**, pp. 80-82.
11. J. Li and L. Zhou, "**Research on Recommendation System of Agricultural Products E-Commerce Platform Based on Hadoop**," *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, 2018, pp. 1070-1073, doi: 10.1109/ICSESS.2018.8663921.
12. P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "**Intelligent Crop Recommendation System using Machine Learning**," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.
13. S. Namani and B. Gonen, "**Smart Agriculture Based on IoT and Cloud Computing**," *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, 2020, pp. 553-556, doi: 0.1109/ICICT50521.2020.00094.
14. GizealewAlazieDagnaw, SisayEbabyeTsigie, "**Applying Logistic Regression Data mining techniques for Ethiopian Government Agricultural Open Data Sets**", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 10, No.2, March - April 2021
15. Joel Mathew Toms, Zebadiah S. Wahlang N., Dr. M. RoshniThanka, Dr. E. Bijolin Edwin, "**Innovative Agricultural Information System with UserFriendly Digital Assistance for Farmers**", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 10, No.2, March - April 2021