An Application of Data Analytics in Agriculture Sector for Multi-Advice Generator in Native Language

Mohit Dalvi	**71710743E**
Samihan Deshmukh	**71710761C**
Devesh Dhannawat	**71710772J**
Pratik Gawali	**71818855B**

Guide - Prof. S. R. Vispute

Contents

Introduction

The main objective of the proposed work is to develop a multi advice generator that will assist farmers in their farming.This system will generate multiple advice in native language of farmers by taking into account various soil conditions and crop peculiarity of the respective area.

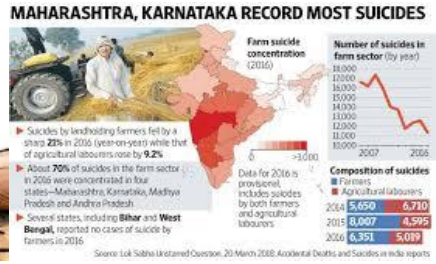
Problem Definition

The project proposes a system to recommend crops, fertilizers, pesticides based on NPK values, pH of soil, water retention capacity, alkalinity, moisture of soil, organic matter content, heavy metal content, crop calendar, weather in that particular region. All advice will be converted to a sentence in native language.

Abstract

Soil is the most important part of agriculture. Different prediction models are developed which use factors which effects on soil health. Soil moisture, nitrogen, phosphorus, potassium is important for farmers to determine how much irrigation is required, which type of crops can be grown in such soil, which fertilizers to use for better yield from soil. This application will be used to generate advices for the farmers related to agriculture using Associative classifier in native language depending upon different factors which affect soil health.

Motivation



The condition of farmers in the State of Maharashtra today is alarming. India is a country where around 50% of the population is dependent on Agriculture as primary source of income. And around 60% to 70% of the population is directly or indirectly dependent on Agriculture as source of income. Despite of this, agriculture has only 17-18% contribution to GDP of India. Main reason behind this is crop failure, less yield of crops, diseases on crops. There are no advance technologies being used in this field and people are still using traditional methods for decision making processes for activities such as crop selection, fertilizers selection, pesticide selection, etc. Farmers fall into debts because they have to face a scarce crop productivity, which increases the risk of their profit. This creates a vicious cycle and most of the time farmers suffer mental distress. The Existing Systems generates only the single advice Hence, we were motivated to help solve farmers Issues through our project.

Objective

To generate dataset in native language and develop multiple advice based on soil health status, NPK values, weather conditions and crop calender using Machine learning algorithms. The advice will be in native language consisting of recommended fertilizers, crops and pesticides .

Literature Review

1) “Soil **Moisture Prediction** Using Machine Learning” by Shikha Prakash et al for predicting soil moisture. They have used machine learning algorithm like multiple linear regression, support vector regression and recurrent neural networks for prediction of soil moisture for 1 day, 2 days and 7 days ahead. The performance was compared with R 2 and RMSE. The comparison shows multiple linear regression is better in providing RMSE and R2 of 0.14 and 0.975 for day1 , 0.353 and 0.939 for day 2, 1.59 and 0.786 for day 7 ahead.

2) “The prediction model for **soil water evaporation** based on Back Propagation neural network” LiliMa et al for predicting soil water evaporation rate. They have used BP neural network for prediction which gives high accuracy and high stability. Average temperature, relative humidity, net radiation these factors are considered for finding the output.

3) “**Soil Fertility** Grading With Bayesian Network Transfer Learning” by Hai-yang Jia et al describes methods for grading soil into different types so that suitable crops can be grown in it. They have used Bayesian Network based transfer learning algorithm. Algorithm includes structural learning parameter learning, considering both similarities between learning task and the geographical position of land square. Empirical experiment results show a significant improvement in terms of structure and parameters when transfer knowledge between similar soil fertility grading task.

4) “Predicting **Soil Heavy Metal** Based On Random Forest Model” by Weibo Ma et al predicts heavy metal contain of soil. They have used three machine learning algorithms SVM, RF, ELM and compared with PLS method. Here **ELM** and **RF** performs better than SVM. The concentration of metals will affect the prediction of ELM. Stability of RF is best among these three models. RF algorithm has higher accuracy for inversion of soil heavy metal research.

5) “Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction” by Jay Gholap et al describes analysis of the soil data using different algorithms and prediction technique. In spite the fact that the least median squares regression is known to produce better results than the classical linear regression technique, from the given set of attributes, the most accurately predicted attribute was “P” (**Phosphorous content** of the soil) and which was determined using the Linear Regression technique in lesser time as compared to Least Median Squares Regression. We have demonstrated a comparative study of various classification algorithms i.e. Naïve Bayes, J48 (C4.5), JRip with the help of data mining tool WEKA. **J48** is very simple classifier to make a decision tree, but it gave the best result in the experiment.

COMPARISON OF SOIL MOISTURE PREDICTION

Model	Soil Depth (in cm)	Accuracy (in %)
SVM	0 to 20	92.899
	20 to 40	92.656
Random Forest	0 to 20	87.632
	20 to 40	87.842
Back-propagation Neural Network	0 to 20	80.570
	20 to 40	85.323

COMPARISON OF SOIL FERTILITY PREDICTION

Objective	Models	Best Model
Predict **Cation Exchange Capacity** using silt, clay and sand contents, OC and pH	GEP, **Next fit (NF)**, NN, SVM	NF
Design PTF to estimate **Soil-index** USING SOC, BD, content of Nitrogen	**ANN**, LR	ANN
Predict **Soil Organic Content** using nine continuous variables and two categorical	MLR, CART, **RF**	RF
Predict **microbial dynamics** using 3 inputs temperature, pH, incubation period	ANN, SVR, WM-FIS, **Subtractive Clustering-Fuzzy. Inference System(SC-FIS)**	SC-FIS
Predict **Soil Water Retention Capacity** using physical and chemical properties	GEP, **NF**, NN SVM, MARS, RF, ML	NF
Produce point Pedotransfer function, PC Pedotransfer function to determine **Soil Water Retention Capacity** using certain physical and chemical properties of the soil	MLR, **ANN**, SVR, kNN	ANN
Predict **Phosphorus content** using clay, sand, organic matter and pH	**ANN**, GA, FIS, ANFIS, PCR	ANN

CADMIUM PREDICTION COMPARISON

Model	Evaluation Indicator	Value
PLS	R^2	0.5362
	RMSE	20.9907
SVM	R^2	0.6025
	RMSE	19.4837
RF	R^2	0.6486
	RMSE	15.5256
ELM	R^2	0.7085
	RMSE	16.6493

ARSENIC PREDICTION COMPARISON

Model	Evaluation Indicator	Value
PLS	R^2	0.9431
	RMSE	1.3683
SVM	R^2	0.9720
	RMSE	0.8760
RF	R^2	0.9912
	RMSE	0.5327
ELM	R^2	0.9653
	RMSE	0.7037

LEAD PREDICTION COMPARISON

Model	Evaluation Indicator	Value
PLS	R^2	0.9071
	RMSE	2.6327
SVM	R^2	0.9622
	RMSE	1.5919
RF	R^2	0.9756
	RMSE	1.1694
ELM	R^2	0.9652
	RMSE	1.6501

COMPARISON OF SOIL CLASSIFICATION ALGORITHM

Model	Accuracy (%)
Gaussian SVM	94.95
Weighted k-NN	92.93
Bagged trees	90.91

COMPARISON OF N, P, K PREDICTION

Element	RMSE
N	0.32
P	0.34
K	0.41

Paper Publication Details

Title : Application of Data Analytics in Agriculture Sector for Soil Health Analysis - Literature Review

Conference : ICCUBEA 2019

Status : Accepted and Presented in ICCUBEA 2019 held at PCCOE Pune.

Paper Publication Details

Title : An Application of Data Analytics in Agriculture Sector for Multi-Advice generator in native language.

Conference : International Journal of Future Generation Communication and Networking, Web of Science (ESCI) .

Status : Accepted.

Copyright Details

Title: MAHARASHTRA DISTRICT WISE CROP DATASET BASED ON SOIL QUALITY AND PH

Diary Number: 5833/2020-CO/L

Status: Re-Scrutiny

Proposed System

1) Mathematical Model:

M = { s, e, X, Y, DD, Fme }

Where,

s = Start state

e = End state

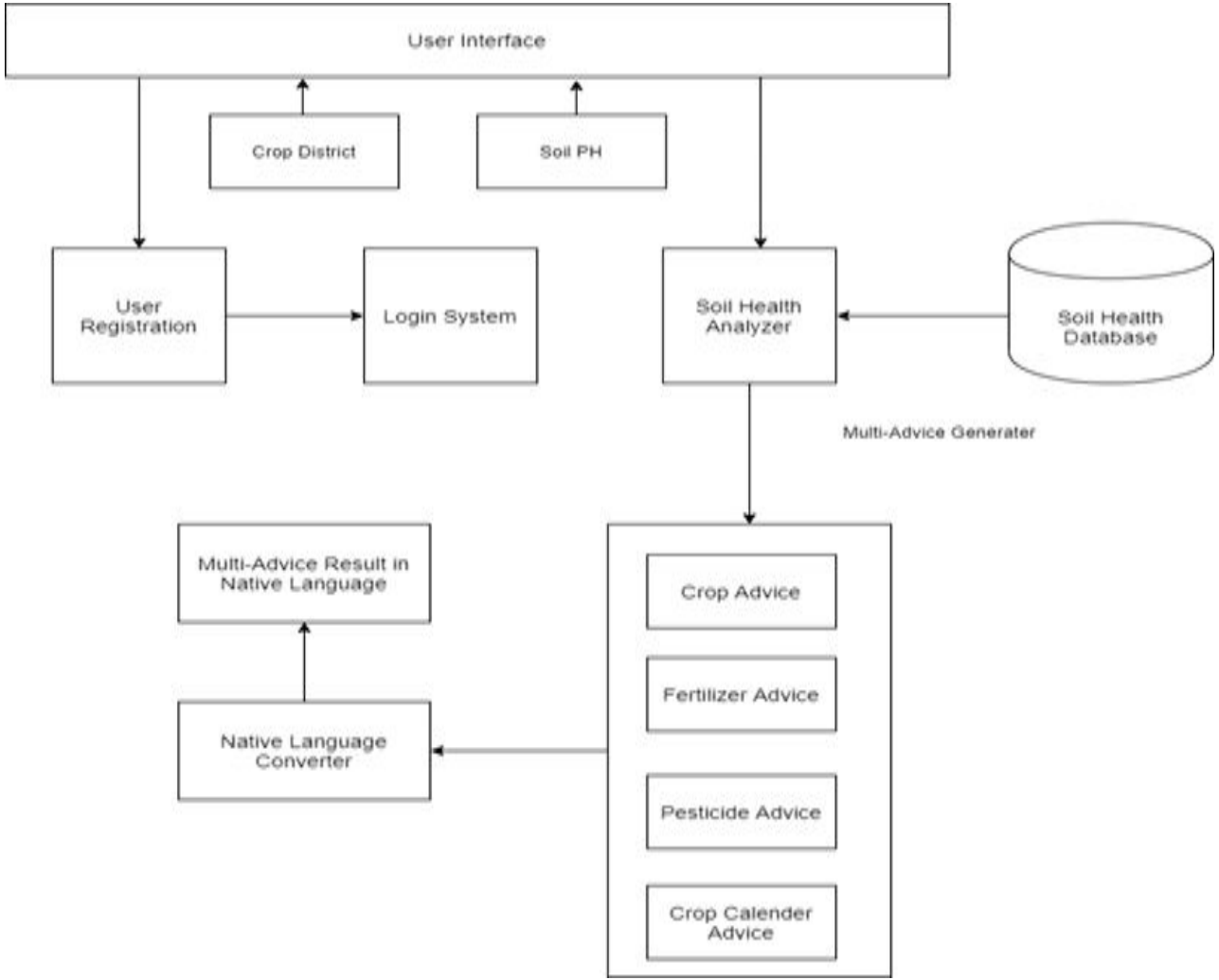
X = input as NPK values, pH, etc factors

Y = output as crops, fertilizers to be used, pesticides, etc.

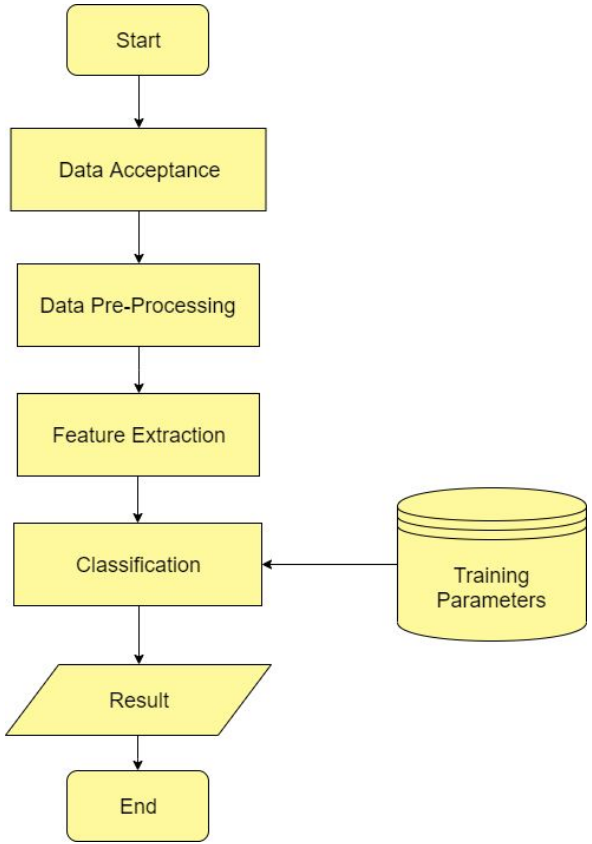
DD = Deterministic data

Fme: Functions used (multiclass_classification_learner(), multiclass_classification_generator())

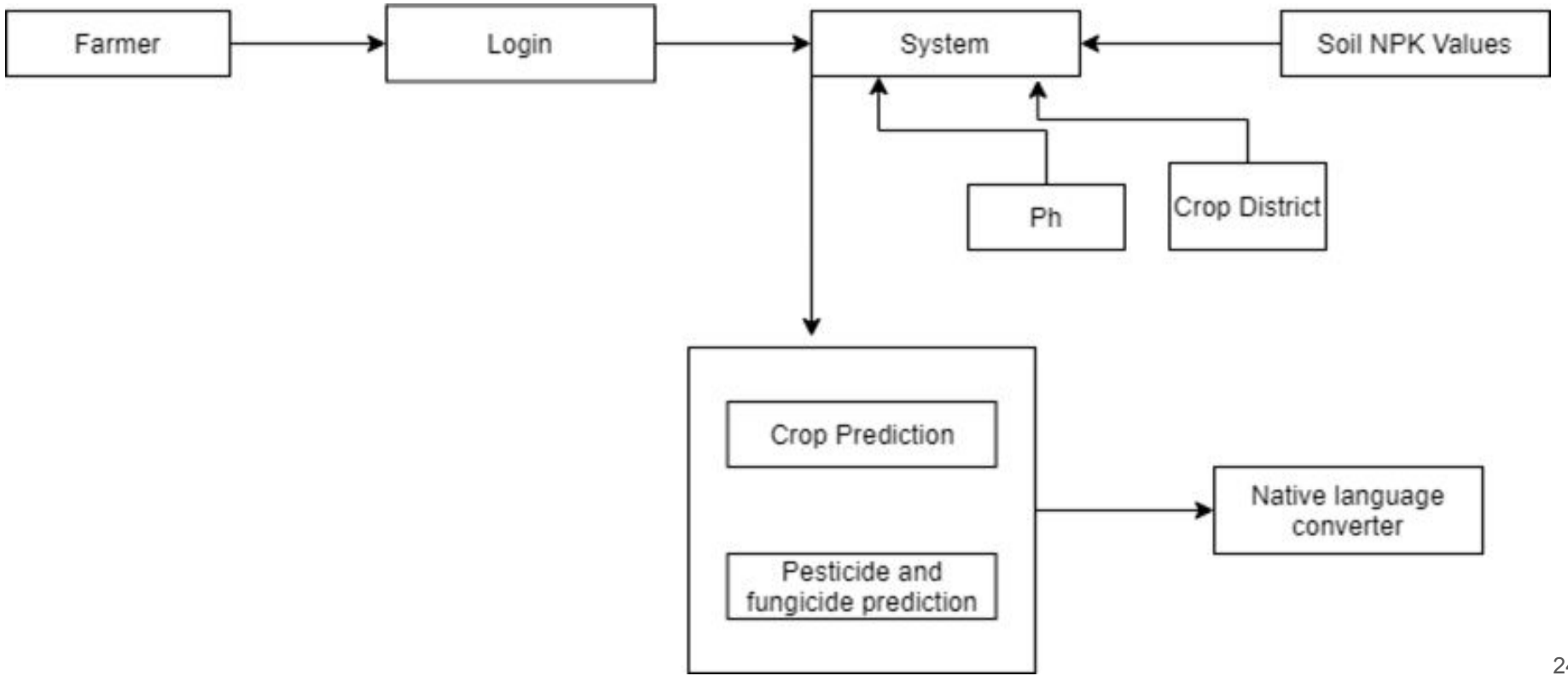
Block Diagram



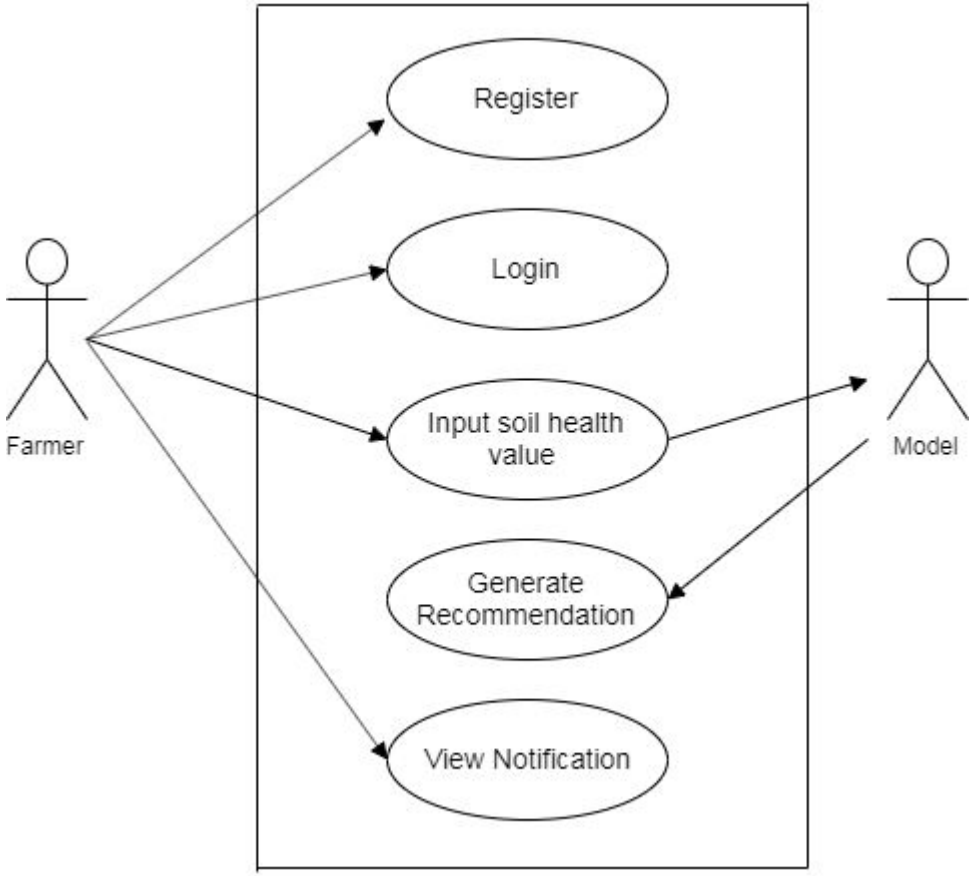
System Implementation Diagram



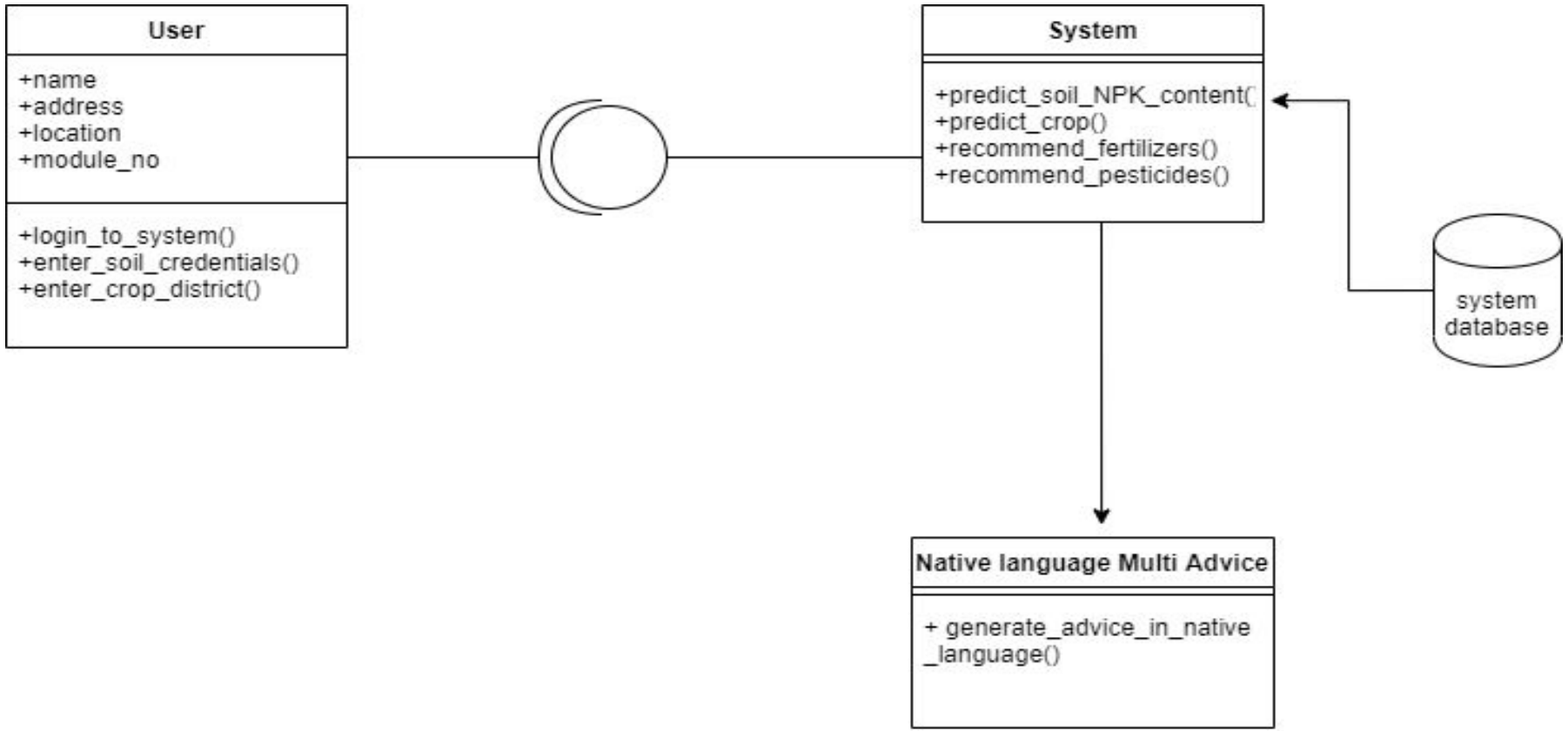
Data Flow Diagram



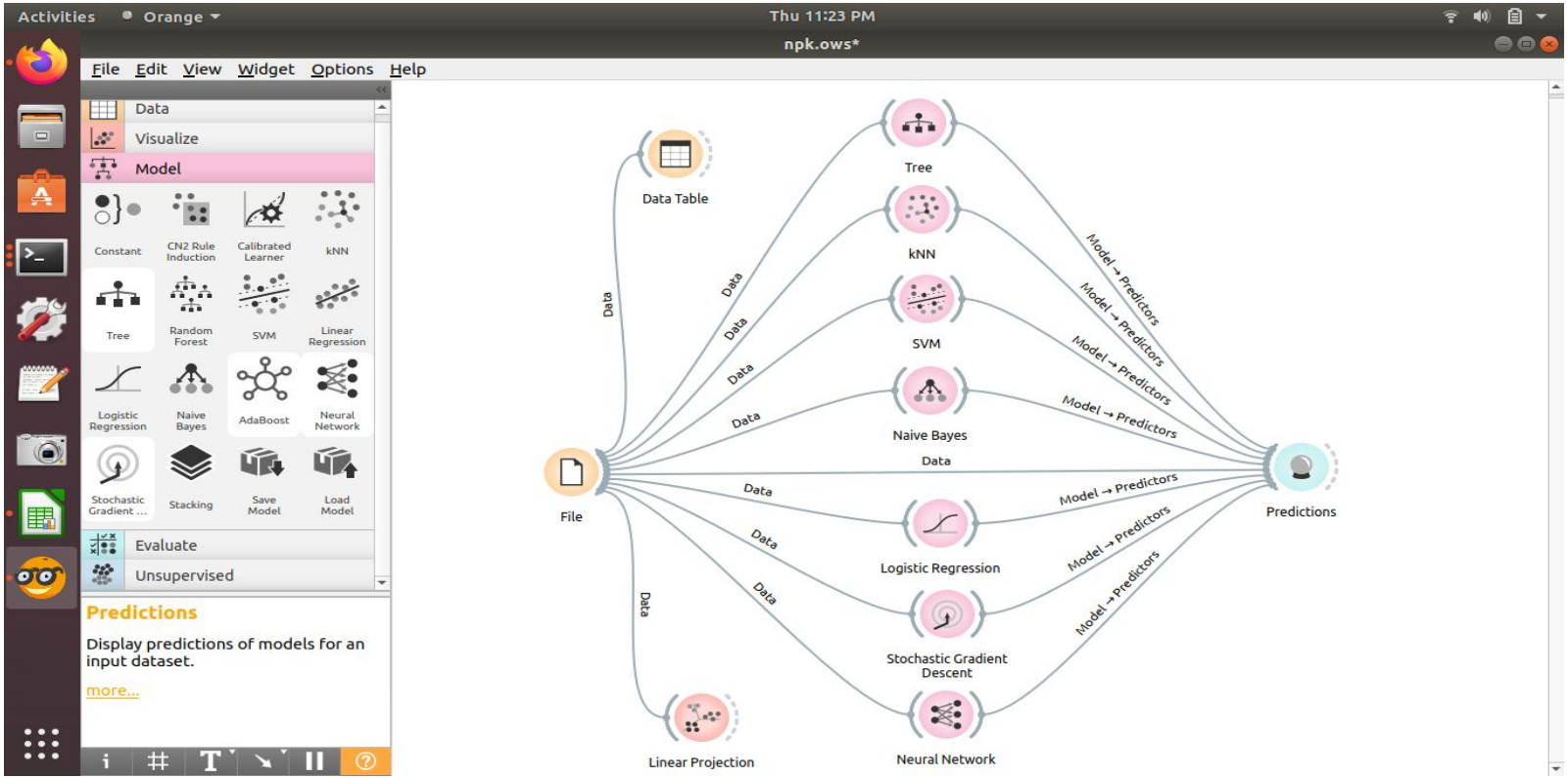
Use Case Diagram



Class Diagram



List of algorithms tried to predict soil N P K values:



CN2 Classifier:

Algorithmic Details

- The representation for rules output by CN2 is an ordered set of if-then rules, also known as a decision list (Rivest, 1987). CN2 uses a heuristic function to terminate search during rule construction, based on an estimate of the noise present in the data. This results in rules that may not classify all the training examples correctly, but that perform well on new data

Decision tree:

Generally, a decision tree comprises of three basic segments including a root node, a few hidden nodes, and a lot of terminal nodes. As demonstrated, for each hidden and terminal node, there should exist a parent node demonstrating the data source. In the interim, with respect to the root node and each hidden node, at least two child nodes will be created from these parent nodes dependent on different decision rules

Language Conversion API

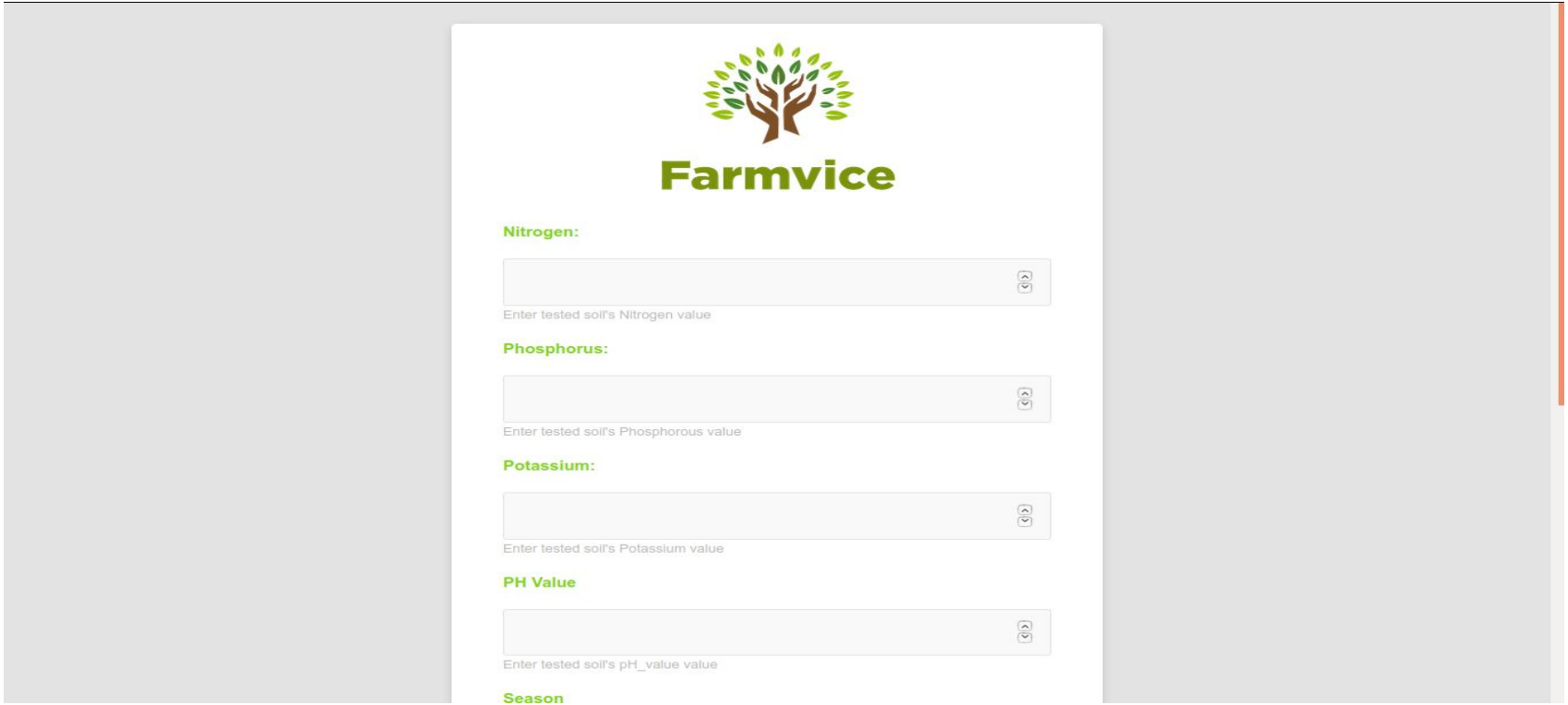
Goslate provides you free python API to google translation service by querying google translation website. It is: Free: get translation through public google web site without fee. Fast: batch, cache and concurrently fetch.

Frontend Technology

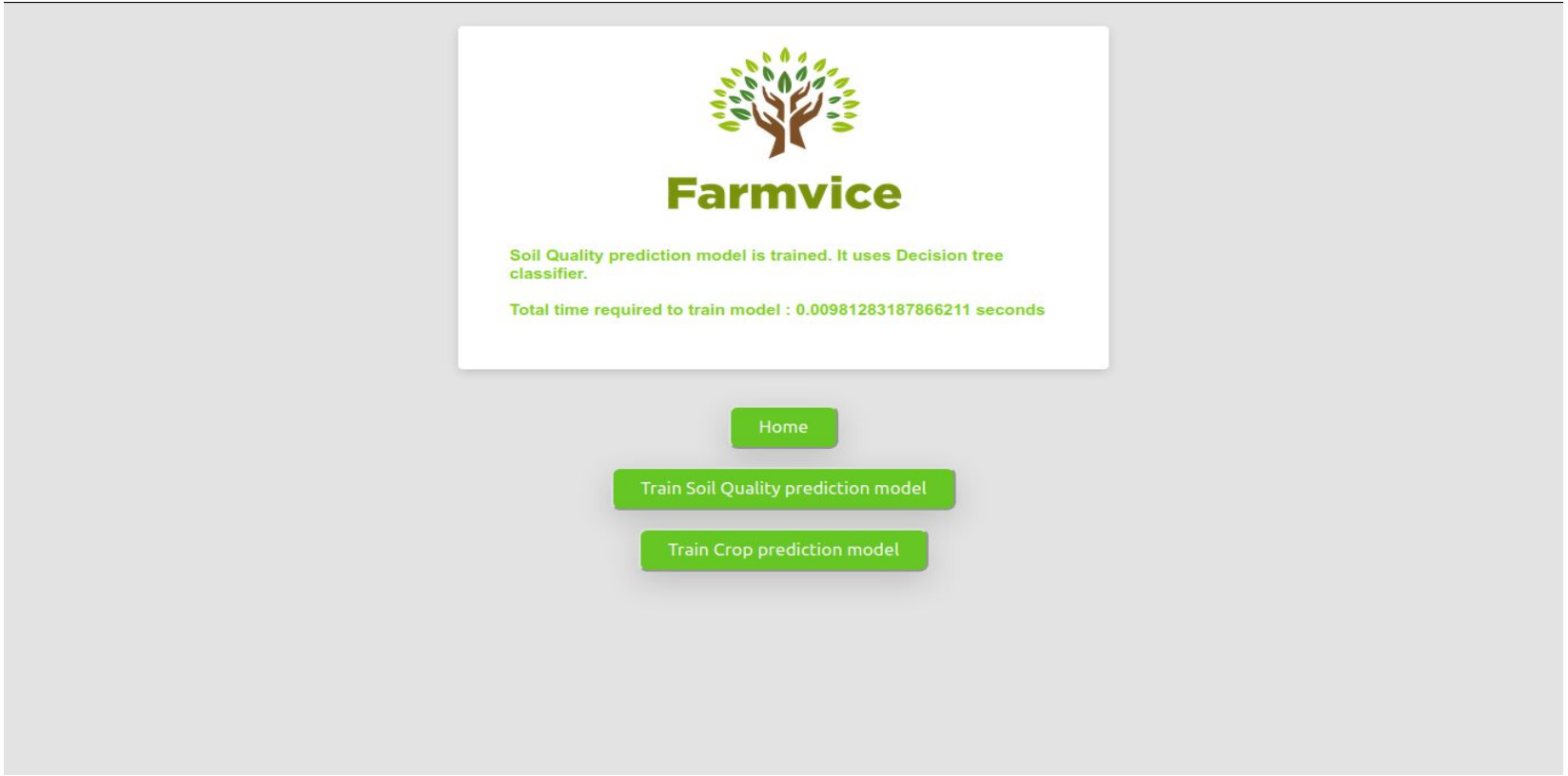
Python Django: Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Screenshots

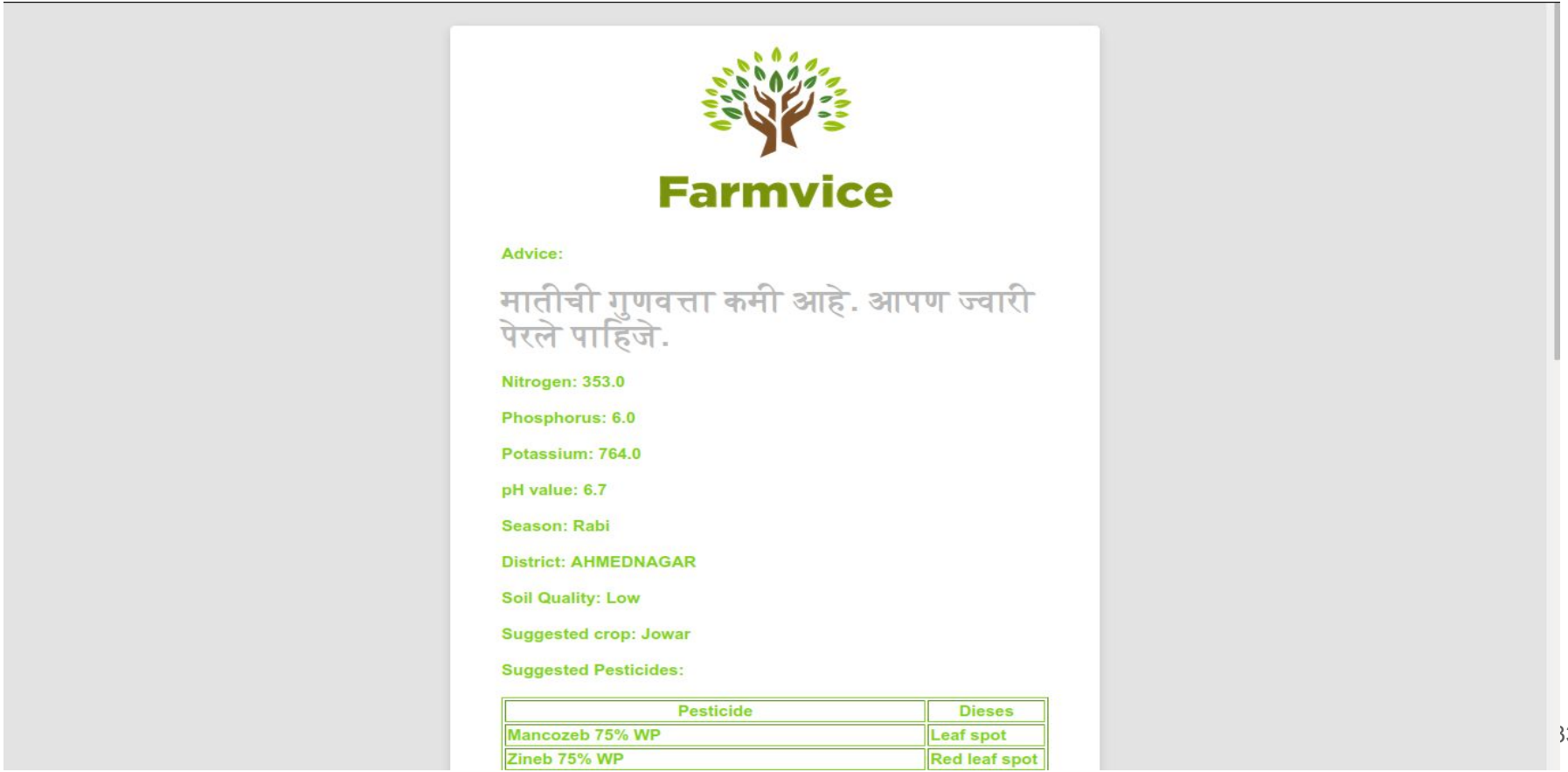
Interface for input parameters



Result after applying Decision Tree algorithm for soil health predicton



Advice generated in native language(marathi) using Goslate



Advice generated suggesting pesticides and fungicides according to the crop

Suggested crop: Jowar

Suggested Pesticides:

Pesticide	Dieses
Mancozeb 75% WP	Leaf spot
Zineb 75% WP	Red leaf spot
Carboxin 17.5% + Thiram 17.5% FF	Leaf spot
Propiconazole 13.9% + Difenoconazole 13.9% EC	Leaf blight

Suggested Insecticides:

Insecticide	Pests
3% Carbofuran, 5% Disulfoton,10% Phorate granule	Shoot fly
apply 4% Endosulfan, carbaryl granules, spray 0.05% Lindane	Stem borers
0.05% Endosulfan, 01% Carbaryl, 0.05% Lindane sprays or dusts	Sorghum midge
0.02% Phosphamidon, 0.04% Diazinon Methyl-demeton , Dimethoate	Aphids
dust hoppers and adults with 5 and 10% BHC, respectively.	Deccan wingless grass hopper
5% BHC or 4% Carbaryl,0.1% Carbaryl	Earhead bug
0.03% Phosphamidon Dimethoate Diazinon, Monocrotophon, Methyl demeton	Sorghum shoot bug
10% BHC	Cut worm
10% BHC	Hairy caterpillars
5% BHC or 4% Carbaryl,spray 0.1% Carbaryl	Earhead cater-pillars
Dust sulphur, spray 0.05% wettable sulphur	Mites

Summary

➢ The implementation of the proposed model would surely be very helpful for the Governmental bodies for giving advice like suggested crop, pesticides, insecticides, fertilizers to the farmers specifically related to the health conditions of their own farmland and increase farmers crop yields and profit.

➢ Model is considering almost all parameters which can affect crop like NPK values, pH of soil, water retention capacity, alkalinity, moisture of soil, organic mater content, heavy metal content, crop calendar, etc.

➢ We used multi-class classifier which increased the accuracy of product. For different components different machine learning algorithms are used which gives the best possible accuracy.

➢ The predicted values are combined to form a sentence. This sentence are in native language which has helped farmers to understand advice of model. Otherwise it could be difficult for farmers to understand advice in English language

References

[1] H. Nie, L. Yang, X. Li, L. Ren, J. Xu and Y. Feng, "Spatial Prediction of Soil Moisture Content in Winter Wheat Based on Machine Learning Model," 2018 26th International Conference on Geoinformatics, Kunming, 2018, pp. 1-6

[2] S. Rose, S. Nickolas and S. Sangeetha, "Machine Learning and Statistical Approaches used in Estimating Parameters that Affect the Soil Fertility Status: A Survey," 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Bangalore, India, 2018, pp. 381-385

[3] W. Ma, K. Tan and P. Du, "Predicting soil heavy metal based on Random Forest model," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 4331-4334

[4] Al Zaminur Rahman, Sk & Chandra Mitra, Kaushik & Islam, S.M.. (2018). Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series. 1-4

[5] Singh, Ravinder & Kaur, Rubleen & Singla, Anshu. (2018). Two phase fuzzy based prediction model to predict Soil nutrient. 80-85

[6] Prakash, Shikha & Sharma, Animesh & Sahu, Sitanshu. (2018). Soil Moisture Prediction Using Machine Learning. 1-6

[7] L. Ma, Y. Wu, J. Ji and C. He, "The Prediction Model for Soil Water Evaporation Based on BP Neural Network," 2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, Changsha, 2011, pp. 641-645

[8] H. Jia, J. Chen, H. Yu and D. Liu, "Soil fertility grading with Bayesian Network transfer learning," 2010 International Conference on Machine Learning and Cybernetics, Qingdao, 2010, pp. 1159-1163

[9] S. R. Vispute and M. A. Potey, "Automatic text categorization of marathi documents using clustering technique," 2013 15th International Conference on Advanced Computing Technologies (ICACT), Rajampet, 2013, pp. 1-5

[10] E. Claverie, J. Lecoeur, V. Letort and P. Cournède, "Modeling soil temperature to predict emergence," 2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), Qingdao, 2016, pp. 28-37

[11] Prachi S., Kirti P., Pooja N., Alisha S., Sushma V. (2018) Automatic Integration and Clustering of Marathi Documents in Different Formats for Effective Information Retrieval. In: Tiwari B., Tiwari V., Das K., Mishra D., Bansal J. (eds) Proceedings of International Conference on Recent Advancement on Computer and Communication. Lecture Notes in Networks and Systems, vol 34. Springer, Singapore

[12] Vinit Savant et al "A Survey On Farmer's Need and Feedback Analysis System" International Journal of Advanced Research in Computer and Communication Engineering, Vol.3, Issue 11, November 2015

[13] R. Vispute, Sushma & Patil, Shrikant & Sangale, Sagar & Padwal, Akshay & Ukarde, Avinash. (2015). Parallel Processing System for Marathi Content Generation. Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015. 575-579. 10.1109/ICCUBEA.2015.118.

[14] Sirsat, Manisha & Cernadas, E & Fernandez-Delgado, M & Barro, S. (2018). Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. Computers and Electronics in Agriculture. 154.120-133.10.1016/j.compag.2018.08.003.

[15] Vispute, S. R., Patil, S., Sangale, S., Padwal, A., & Ukarde, A. (2015). Parallel Processing System for Marathi Content Generation. 2015 International Conference on Computing Communication Control and Automation.