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

Pratik Gawali[1], Mohit Dalvi[2], Devesh Dhannawat[3] , Samihan Deshmukh[4], S. R. Vispute[5]

[1]*Undergraduate Student, Pimpri Chinchwad College of Engineering, Pune*
pratikgawali06@gmail.com
[2]*Undergraduate Student, Pimpri Chinchwad College of Engineering, Pune*
dalvimohit@gmail.com
[3]*Undergraduate Student, Pimpri Chinchwad College of Engineering, Pune*
devesh.dhannawat@gmail.com
[4]*Undergraduate Student, Pimpri Chinchwad College of Engineering, Pune*
samihan25@gmail.com
[5]*Professor, Pimpri Chinchwad College of Engineering, Pune*
visputesushma@gmail.com

----------***----------

Abstract - *Farmer suicide is one of the major concerns of India.Around 60% to 70% of the population is directly or indirectly dependent on Agriculture as their source of income.Most of the farmers in India fail to produce suffiecient yield due to various uncertainities and problems faced in agriculture.This is the reason farmers take loans,fall into debts and end up giving up their lives.Technological advancements have not been able to make a significant impact in guiding farmers in improving soil health as per crop requirements,crop selection, fertilizers selection, pesticide selection, etc.in order to ensure proper yield.Also,the current systems generate advice based only on single feature of agriculture.To overcome the limitations of the presesnt system, this paper proposes a multi advice generator that takes into account various soil health parameters and crop district based on which multiple advice related to soil health,fertilizer and pesticide selection for respective crops is provided to farmers in native language.Various machine learning algorithms are studied and implemented for accurate soil health and advice generation.*

Key Words: Decision Tree, CN2 Algorithm, Multi Advice generator, Soil quality

1.INTRODUCTION

Majority of Indian people have agriculture as their primary occupation. Around 70% population in India is directly or indirectly dependent on agriculture for their livelihood. India's GDP from agriculture averaged upto 4191.03 INR Billion from 2011 until 2019, reaching an all time high of 6091.05 INR Billion in the fourth quarter of 2019.

There are various technological advances coming up in agriculture but a very few have been able to make a significant impact to farming in order to assure proper crop yield.A major problem faced by Indian farmers is lack of proper assistance in farming that is customized to the soil and climatic conditions of their specific locality.Many farmers in India fail to generate sufficient crop produce due to the lack of timely guidance and various uncertainties which arise during the course of farming.Various systems have been made which give farmers advice based only on one particular feature.We have made an attempt to take into account various features and develop a multiple advice generator to assist farmers that would have a significant impact to the crop produce. Agriculture majorly depends on 2 factors, one is soil fertility and the other is climatic conditions. Our system takes into account various soil conditions like the NPK values,soil pH,district and the crop calendar.Based on this the system generates multiple advice in farmers' native language suggesting them deficient amount of NPK content in soil and the usage of pesticides and fungicides that will be suitable for their crops.

The Existing Systems generates only the single advice dependent on a single feature of agriculture.To overcome the limitations of current systems and to help farmers we have proposed a multi-advice generator that will generate various advice in farmers' native language by taking into account multiple soil health features affecting crop produce.

2. DATASET DESCRIPTION

We built our crop dataset by using different sources. We got a NPK values dataset from the Agricultural department of maharashtra state which has different values of N(Nitrogen), P(Phosphorus), K(Potassium) and respective soil quality values. We got district wise crop yield data from data.org which contains district wise crop yield in the respective district like which different crops are grown most in that respective district. The range of pH of soil in which the respective crops can grow was added to the dataset from using references from different websites like the website of many agricultural University websites and other trusted agricultural websites. The final dataset formed after combining all these datasets contains columns soil quality, pH range of soil, district and crop to be planted. The way in which we can use this dataset is first we can predict soil quality using npk values and then then use soil quality,pH and district to predict the crop.

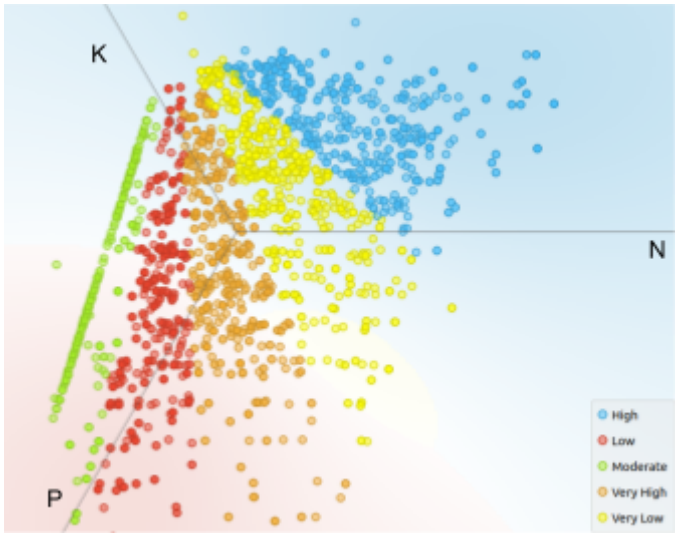

Chart -1: N,P,K value distribution and Soil quality

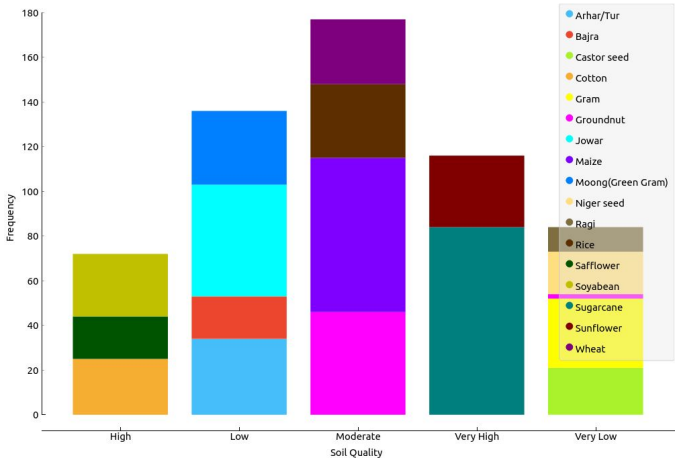

Chart -2: Crop frequency vs Soil quality

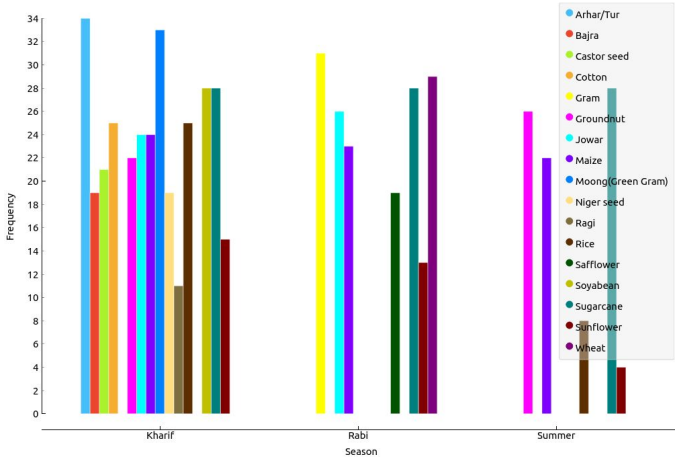

Chart -3: Crop frequency vs Season

3. ALGORITHM DETAILS

We are using two algorithms. one is the Decision tree Algorithm and other is the CN2 Algorithm. Both algorithm are explained in short below:

3.1 Decision Tree

Decision tree is one of the supervised machine learning techniques. The goal of decision tree is to generate a simple decision rule tree to predict the target variable class. These rules are developed based on the training dataset provided. For predicting each class it always starts from the root node and according to the conditions at each node further node is chosen. When it reaches the leaf node it stops and one class for the target variable is decided.

Table -1: Soil quality prediction model comparison

Model Name (Soil NPK Prediction)	Accuracy
Decision Tree	0.984
K Nearest Neighbors	0.959
SVM (Support Vector Machine)	0.912
Naive Bayes	0.695
Logistic Regression	0.876
Stochastic Gradient Descent	0.629

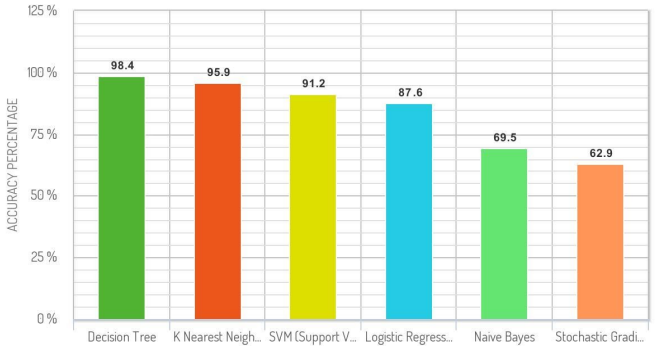

Chart -4: Soil quality prediction model comparison graph

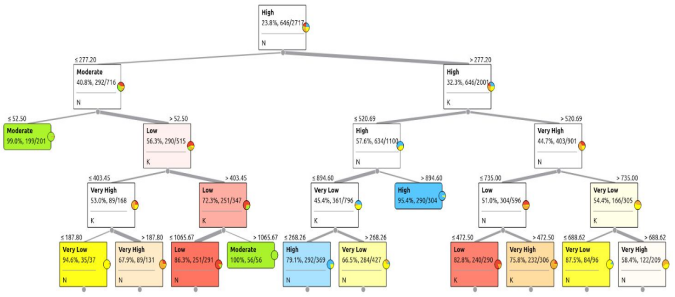

Fig -1: Decision tree of Soil quality prediction with NPK

3.2 CN2 Algorithm

CN2 Algorithm induces rules from given data. This algorithm generates simple, comprehensible if else type rules. Speciality of this algorithm is it can be used in domains where noisy data is present. It works very well even if data is noisy to some extent. The CN2 algorithm can be used only for classification purpose. It is one of the supervised machine learning algorithms.

Table -2: Crop prediction model comparison

Model Name (Crop Advice Generation)	Accuracy
CN2 Rule Inducer	0.94
Adaboost	0.89
KNN (K-Nearest Neighbors)	0.82
Random Forest	0.85
SVM(Support Vector Machine)	0.80
Logistic Regression	0.79
Naive Bayes	0.81
SGD (Stochastic Gradient Descent)	0.86

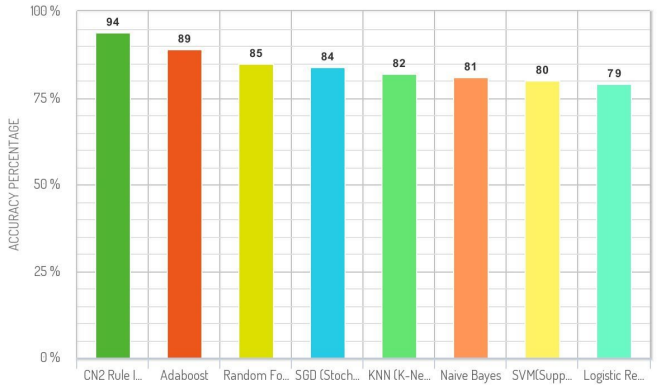

Chart -5: Crop prediction model comparison graph

4. METHODOLOGY

The below Architecture diagram describes our model very well. The user can interact with the system through an interface, the inputs (District, NPK values, PH value) will be provided by the user. The given inputs will be analyzed by the smart algorithm using Soil Health dataset. The trained model will generate multiple advices such as Crop Advice, Fertilizer Advice and Soil Health. Additionally, according to crop advice the algorithm will suggest which pesticides to be used for particular crop disease. These multiple advices will be generated and converted into native language that the user understands.

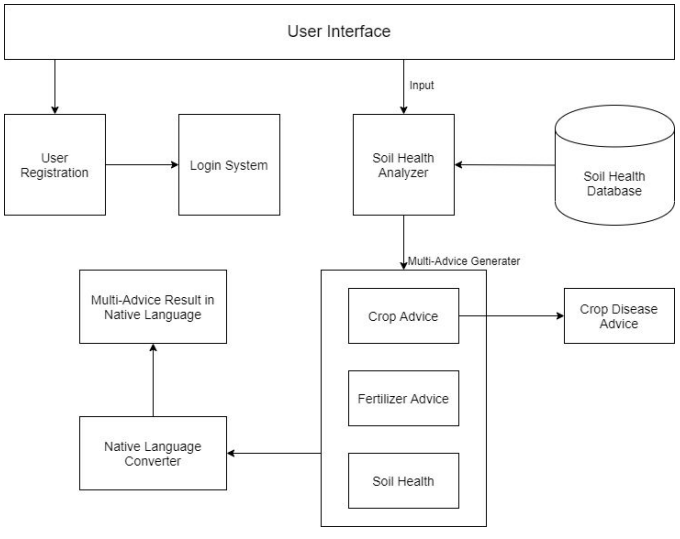

Fig -4: Architecture diagram

The user can interact with the system through an interface, the inputs (District, NPK values, PH value) will be provided by the user. The model will be trained using a training dataset. The given inputs will be analyzed by the smart algorithm using Soil Health dataset. The trained model will generate multiple advice such as Crop Advice, Fertilizer Advice and Soil Health. Additionally, according to crop advice the algorithm will suggest which pesticides to be used for particular crop disease. These multiple advice will be generated and converted into native language (using the goslate library of python) that the user understands.

5. CONCLUSIONS AND FUTURE WORK

In our work, we have attempted to develop a multi-advice generator in native language with a motive to help farmers in agriculture in order to ensure healthy crop production. We have studied and experimented with various machine learning techniques to achieve maximum possible accuracy in predictions based on crop and soil health. Amongst a variety of algorithms tested, Decision Tree Classifier predicts soil health-nitrogen, phosphorus and potassium content with highest accuracy of 98.4 %.These soil health parameters are then fed as input for predicting crops. Experiments show that CN2 predicts crops with highest accuracy of 94 % which is significantly more than other algorithms tested for crop prediction.

This project can be further developed to generate more valuable advice by taking into account various other parameters affecting crop health.For instance,Real time weather conditions can be taken into account along with present parameters while making predictions and generating advice.The more parameters influencing yield will be considered for predictions,the more real and correct will be the advices generated. Apart from this, initially the system makes predictions based on a limited dataset but with successive predictions the system learns from input making the dataset richer resulting in increased accuracy in predictions with time.

REFERENCES

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

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

[3] 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.

[4] 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

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

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

[7] 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

[8] 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.

[9] 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

[10] 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

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

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