Crop Recommendation System

Abstract

Agriculture is the foremost means of livelihood of the majority of the population even to this day. Nearly 50% of the population in India depends on agriculture for their livelihood. But unfortunately, its contribution to the GDP of the country is roughly around 18%. This is mainly because of lack of technology in the agricultural sector. Our farmers still use traditional methods of farming which is a major hindrance for their scalability. Our farmers commit a lot of errors while choosing the type of crop to be grown in the given piece of land. This is majorly affecting the yield of the farmer. This inturn degrades the soil quality too. So we propose here a Crop Recommender System which recommends the type of crop to be grown by the farmer considering the soil type and climatic conditions of the place. We have used Classification algorithms to determine the optimal crop to be grown. This includes Random Forest, Support Vector Machine (SVM), Decision Tree, Artificial Neural Network (ANN), K-nearest neighbor (KNN) and Multivariate Linear Regression (MLR). However Random Forest gave the highest accuracy of 99.01%.

Keywords

Crop Recommendation System, Machine Learning, Random Forest, Support Vector Machine (SVM), Decision Tree, Artificial Neural Network (ANN), K-nearest neighbor (KNN) and Multivariate Linear Regression (MLR), Data Visualization, Outlier Management

Introduction

Farmers completely do not understand the nature of the soil and climate they are growing crops in. This then leads to wrong crop selection. Growing the wrong crop results in the degradation of soil quality and lower crop yield.

The data used for this analysis is a public dataset hosted by open Data by Kaggle. Various data cleaning steps were performed to work on a tidy dataset. After calculations, graphs were plotted to visualize the results and concluded.

Machine learning models are vastly used in predictive analysis due to their ability of making the system learn without explicitly performing the instructions. Previous data is used to train the model. All the attributes, qualitative or quantitative can be used to train the model.

We studied the general crop and soil trends, regarding the amount of nitrogen, phosphorus, potassium, etc. to do extensive research, by factoring in every aspect we deemed important. The dataset was first taken from an opensource site, cleaned, imputed and then worked on.

The dataset was judged on several aspects, both qualitative and quantitative. This analysis provides insights in observing the trend of crops. It shows the most profitable crop based on the parameters entered by the farmer.

Literature review

This paper [1] has proposed to build a web application that can predict the crop yield of a particular crop in each area and climate. They have compared various algorithms for their accuracy. And it is found that Random Forest has given the highest accuracy of 95%.

This paper [2] proposes to build a system that consists of 2 sub-systems. One subsystem will train a model for soil and environmental conditions. The other subsystem will train a model for precipitation parameters. They have also included a Map visualization feature in this model. They have also used multiple machine learning and deep learning models. In that, it is found that neural network gives the highest accuracy of 91%.

This paper [3] proposes an ensemble model in which they are using different classification models simultaneously to derive the best output. The algorithms used here include Random Tree, Chaid, KNN, and Naïve Bayes.

This paper [4] proposed a system of IoT and ML is enabled for soil testing using the sensors, which helps in measuring and observing different parameters like temperature, humidity, moisture content etc. This system lowers the probability of soil degradation and helps maintain crop health.

This paper [5] aims to deliver direct advisory services to even the smallest farmer at the level of his/her smallest plot of crop using the most accessible technologies such as SMS and email. Proposes a Crop Selection Method (CSM) that solves the crop selection problem and improves net yield. The most relevant and accurate classifiers is identified.

The main approaches discussed in this research paper [6] are precision farming techniques for soil fertility monitoring, variable rate application (VRA) of fertilizers or other agrochemicals, remote sensing data analysis for yield prediction models and automation through robotics and machine learning algorithms applied to decision support systems.

The key takeaways from this research paper [7] are that precision agriculture can help address the common problem faced by Indian farmers of not opting for crops based on soil necessity, resulting in low productivity.

Soil testing labs take a considerable amount of time in providing the results of the soil samples. This system [8] claims that it helps the farmers to get a better crop prediction. The proposed method efficiently estimates the soil nutrients based on the data fetched by the sensor network.

Precision agriculture has shown an improvement with time but there are still some issues which need to be addressed. This research [9] is proposing a system where the major factors are taken into consideration at the same time and come up with a solution so that the system will not be complicated for the user.

Methodology

The solutions that exist in the market as of now lack accuracy and they fail to take all possible factors into account. Therefore we decided to rest our focus on improving the accuracy of our model and also focussing on all possible parameters.

We collected our dataset from kaggle named "Crop Recommender System". We took into consideration the Nitrogen, Phosphorus, Potassium content and the pH value of the soil. It also took into account the climate parameters like temperature, rainfall and humidity. Considering all these factors the dataset gave the optimal crop that should be grown in the given piece of land. The dataset consists of 22 different crops. It has a total of 2200 observations.

Before moving directly into the implementation of the algorithms on the dataset, it is necessary for us to first understand the structure of the dataset given to us. Therefore we decided to visualize the given dataset to get clarity about the algorithms we must use here.

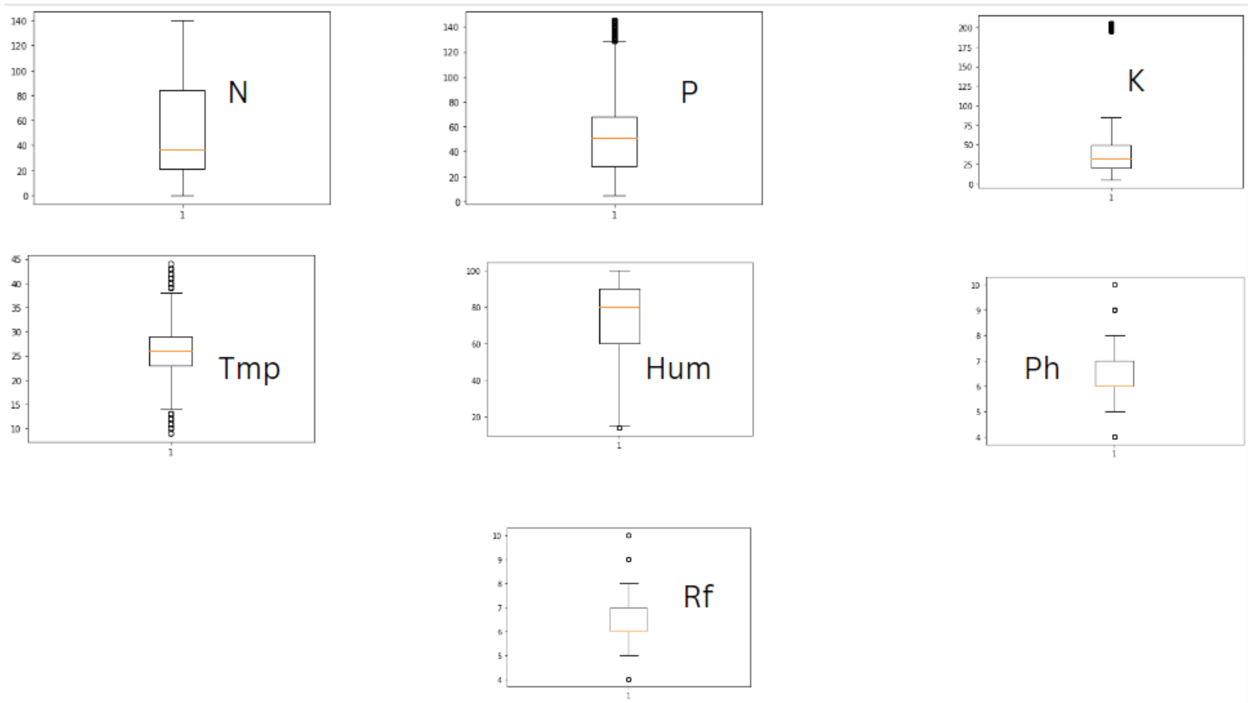

Fig. 1 Box Plots

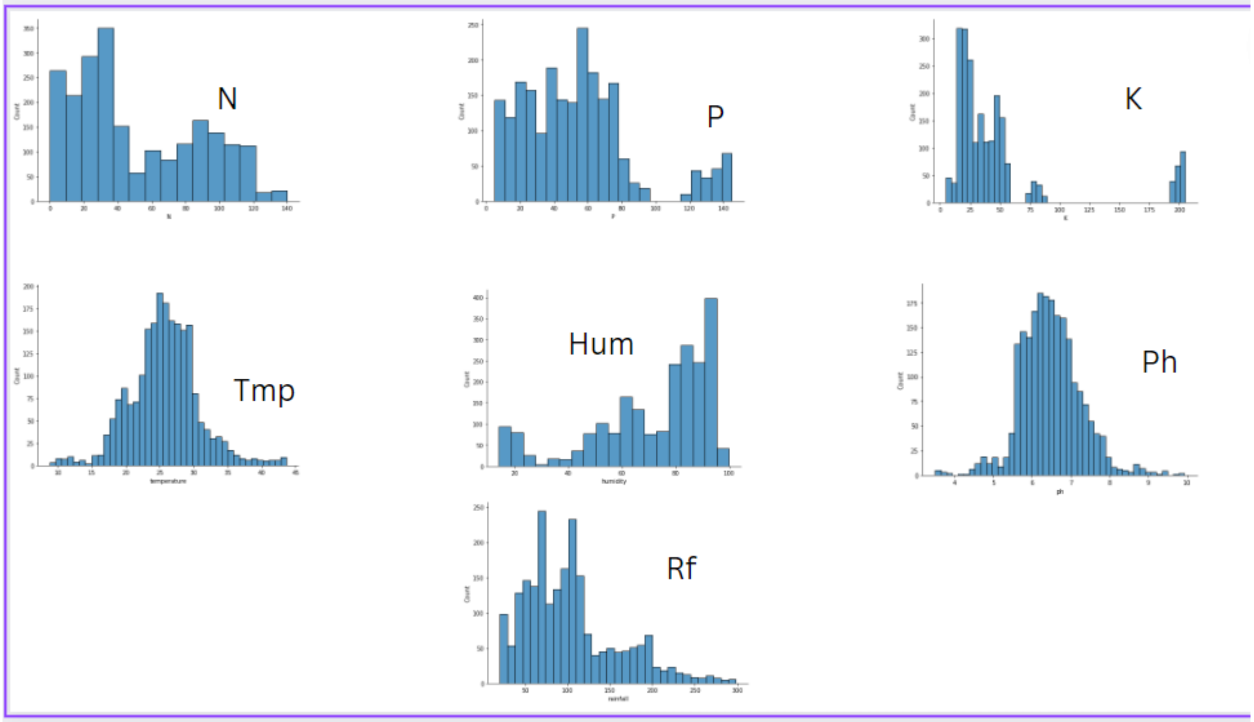

Fig. 2 Frequency distribution

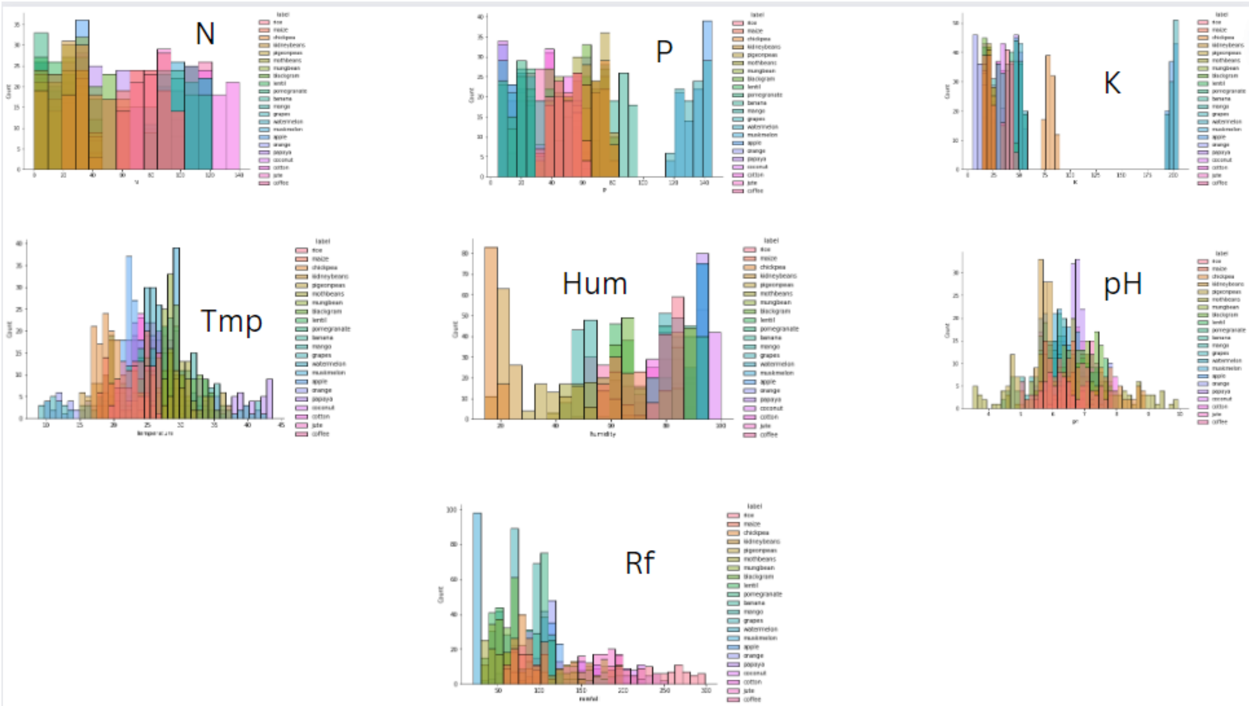

Fig. 3 Frequency distribution based on crops

Fig. 1 depicts the box plots of different parameters in the dataset(N- nitrogen, P- phosphorus, K- potassium, Tmp- temperature, Hum- humidity, pH- pH value of soil, Rf- rainfall). The box plot gives the range of values for a specific parameter.

Fig. 2 depicts frequency distribution of values of each parameter. This visualization gives us the information about how the values are distributed for a given parameter. We can infer from the graphs that Temperature and pH parameters are normally distributed while other parameters are not distributed normally.

Fig. 3 is just the modification of frequency distribution of each parameter. The update here is that different colors are assigned for labels(crops). This will give us the frequency distribution of each crop in our dataset.

From the visualizations depicted above, we can infer that-

- The data needs to be normalized.
- Outliers need to be removed.

Therefore we standardize the dataset using python built-in libraries. We use the StandardScalar library to do this. Then we have to do outlier removal. We have used the IQR method for this. We cannot use methods like the z-score because our data is not normally distributed.

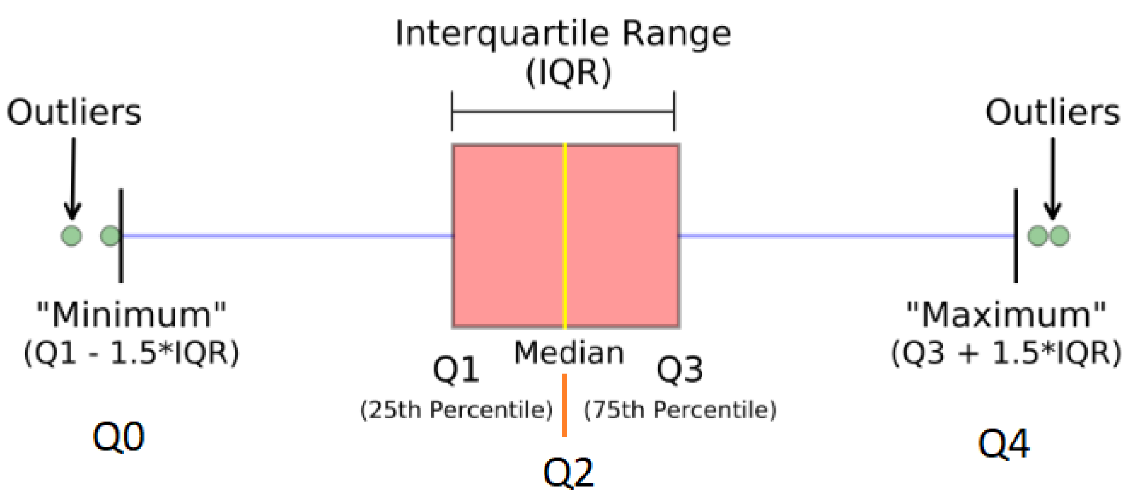

Fig. 4 IQR Method

After removal of outliers, we are left with 2028 data points as compared to 2200 data points previously. Now we have our dataset ready for model creation.

We have applied several classification algorithms to our dataset. They include-

- Decision Tree
- KNN
- Random Forest
- SVM
- Multivariate Linear Regression

- ANN

Most algorithms gave us accuracy above 90%. Random Forest gave us the highest accuracy of 99.014%. Therefore we decided to use Random Forest to build our model.

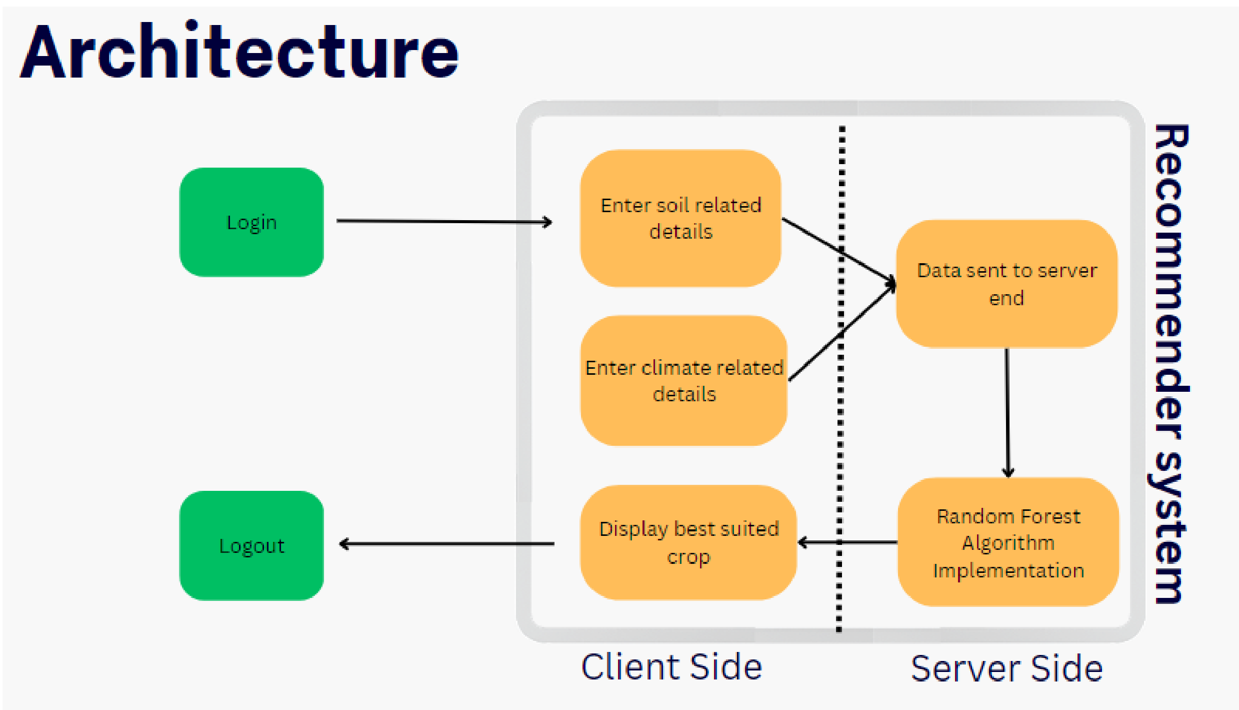

Fig. 5 System Architecture

Fig. 5 depicts the system architecture we are proposing. It has a login and logout option. It collects soil and climate related information from the user. At the backend it will run the Random Forest Algorithm to determine the best suitable crop. Then finally it displays the best suitable crop on the screen.

Results

Algorithms and Accuracy

Algorithm		Accuracy
Random Forest		99.01%
SVM	Linear	98.27%
	Polynomial	93.59%
	Gaussian	97.78%
	Sigmoid	88.66%
KNN		91.13%
Decision Tree	Gini	17.95%
	Entropy	30%
ANN		94.7%
Multivariable Linear Regression		R^2 score = 0.25

Fig. 6 Algorithm Accuracy Comparison

Fig. 6 gives accuracy comparison between different classification algorithms we have used in this study. We observe that Random Forest gives the highest accuracy followed by SVM and KNN.

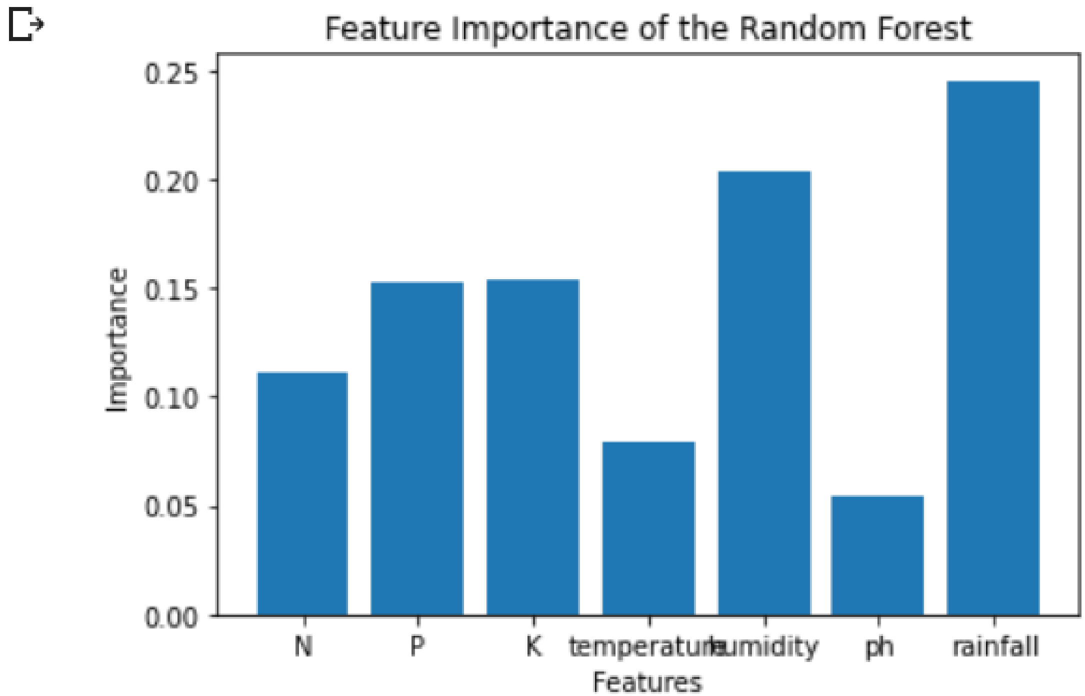

Fig. 7 Feature Importance Plot

Since Random Forest gives the maximum accuracy, we decided to use that algorithm for training our model. Fig. 7 gives the feature importance when the model is trained using the Random Forest algorithm. We can infer from the plot that Rainfall is the most important feature followed by humidity.

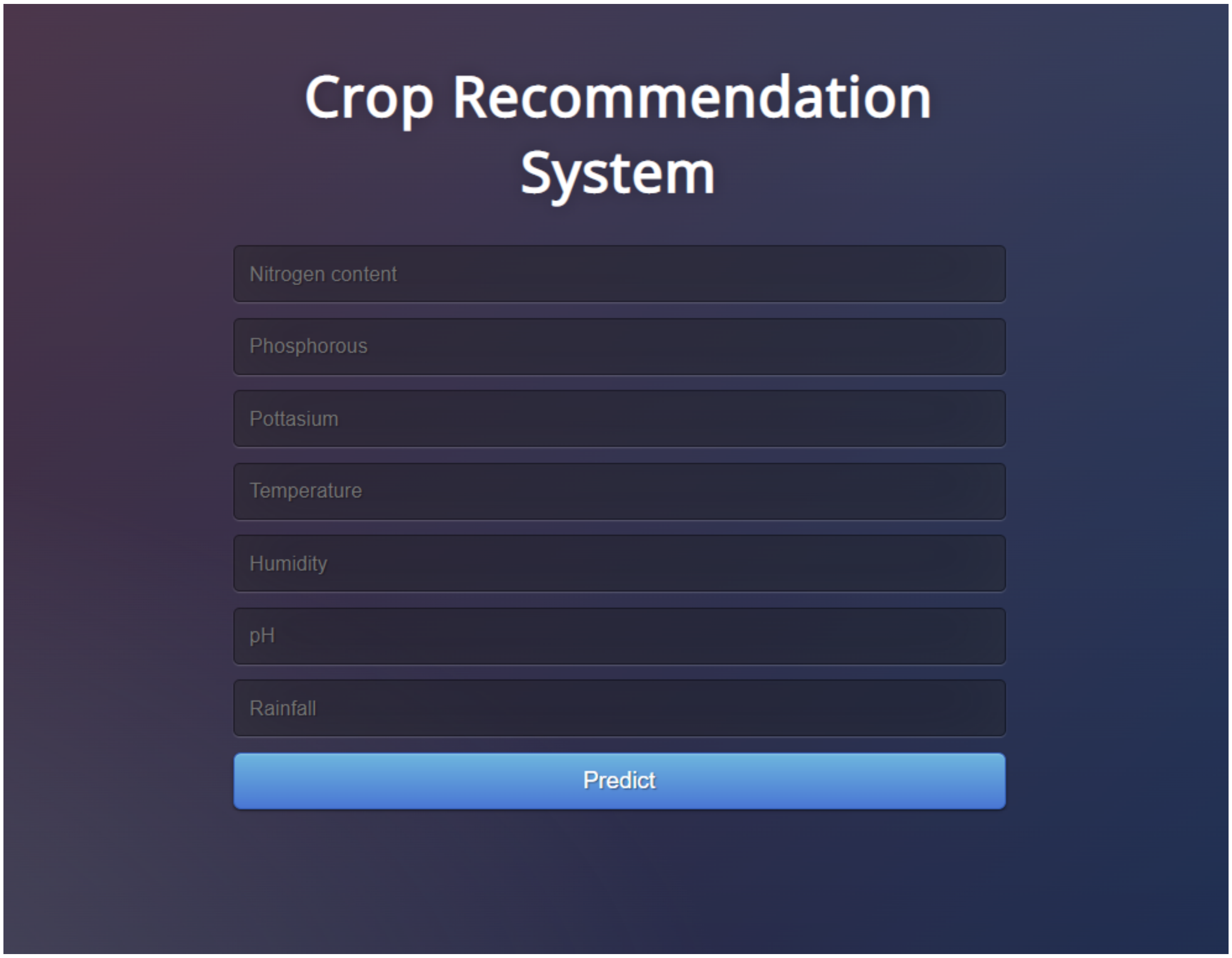

Fig. 8 UI of Our System

Fig. 8 shows the User Interface of our Recommendation System. This takes the input from the user regarding soil and climatic conditions. Based on the information given by the user, it gives the optimal crop that can be grown.

Conclusion

In order to assist farmers in choosing the ideal crop based on the needs of the soil, this research suggests a workable recommendation system. To compare their accuracy, a variety of machine learning methods including Random Forest, ANN, SVM, MLR, and KNN were implemented. With a 99.01% accuracy, the results indicated that Random Forest Regression was the best. The model can be further optimized by involving economic parameters. An enhanced data collection with a high number of features and yield prediction are the goals of future development.

Acknowledgement

We wish to express our sincere thanks and deep sense of gratitude to our guide, Dr. Vinothini A., Vellore Institute of Technology, for her consistent encouragement and valuable guidance offered to us in a pleasant manner throughout the course of this work.
We are extremely grateful to Dr. Susan Elias, Dean of School of Electronics Engineering, VIT, for extending the facilities of the school towards our work and for her unstinting support.
We also take this opportunity to thank all the faculty of the school for their support and their wisdom imparted to us throughout the course.

References

1. Pande, Shilpa Mangesh, et al. "Crop recommender system using machine learning approach." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.

2. Lakshmi, N., et al. "Crop recommendation system for precision agriculture." International Journal for Research in Applied Science and Engineering Technology 6.5 (2018): 1132-1136

3. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.

4. Pande, S. M., Ramesh, P. K., ANMOL, A., Aishwarya, B. R., ROHILLA, K., & SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1066-1071). IEEE.

5. Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.

6. Mondal, P., & Basu, M. (2009). Adoption of precision agriculture technologies in India and in some developing countries: Scope, present status and strategies. Progress in Natural Science, 19(6), 659-666.

7. Bhattacharyay, D., Maitra, S., Pine, S., Shankar, T., & Pedda Ghouse Peera, S. K. (2020). Future of precision agriculture in India. Protected Cultivation and Smart Agriculture, 1, 289- 299.

8. Singh, A. K. (2022). Precision Agriculture in India–Opportunities and Challenges. Indian Journal of Fertilisers, 18(4), 308-331.

9. Pinaki, M., Basu, M., & Bhadoria, P. B. S. (2011). Critical review of precision agriculture technologies and its scope of adoption in India. American Journal of Experimental Agriculture, 1(3), 49-68.

10. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision

agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.