Grenze International Journal of Engineering and Technology, June Issue

Soil Analysis and Crop Prediction

Mr. Hirdesh Sharma[1], Rishabh Gosain[2], Prerna Sachdeva[3], Sahej Budhiraja[4], Ridhima Chhabra[5], Mr. Ravi Chaturvedi[6] and Ms. Annu Mishra[7]
[1]Assistant Professor, Dept. of CSE, JIMS Engineering Management Technical Campus, (JEMTEC), Greater Noida, U.P., India, ORCID ID: 0000-0002-1278-4135
Email: hirdesharma@gmail.com
[2-5]JIMS Engineering Management Technical Campus (JEMTEC), Greater Noida, U.P., India
Email: rishabhgosain26@gmail.com, prernasachdeva68@gmail.com, sehajbudhiraja2000@gmail.com, ridhimachhabra12@gmail.com
[6-7]Assistant Professor, Dept. of CSE, JIMS Engineering Management Technical Campus (JEMTEC), Greater Noida, U.P., India
Email: rpchaturvedi51@gmail.com, pandey_annu18@yahoo.com

***Abstract*—To find out what plant nutrients are in the soil that are available, soil analysis is a crucial procedure. The main source of nutrients for plants is the soil. Apart from soil, there are other important variables that impact plant development, such as fertilizer, rainfall, and precipitation. Our goal is to develop a prediction engine that can identify which crop would be best for a given soil. Soil quality analysis is the process of examining the characteristics of the soil to ascertain if it is appropriate for fostering plant development. It entails evaluating a variety of factors, such as pH level, nutritional content, organic matter concentration, and texture. This assessment helps identify any deficiencies or imbalances in the soil that can affect crop yield. Crop suggestion is the process of suggesting suitable crops to cultivate based on the results of soil analysis and other relevant factors. There are three characteristics that are used to classify the kind of soil. Soil moisture, methane, and nitrogen make up the three characteristics.**

***Index Terms*— Nitrogen, Phosphorus, Potassium, Random Forest, Crop Prediction, Algorithms.**

I. INTRODUCTION

In order to maximise agricultural productivity and sustainability, crop recommendations and soil quality analyses are crucial. We can assess the fertility, nutrient availability, pH level, and overall health of soil by examining its physical, chemical, and biological properties. When selecting crops, the soil quality analysis findings are considered in addition to climate, availability of water, and market demand. Crops can be matched to the unique properties of the soil to maximise yields and save expenses. It enhances the long-term productivity and profitability of farming companies, reduces its detrimental impacts on the environment, and promotes the effective use of resources. By evaluating the condition of their soil and carefully choosing their crops, farmers may increase yields, improve the health of the soil, and contribute to food safety and sustainability of the environment, and all of these benefits.

II. LITERATURE SURVEY

Research on the agricultural industry has been conducted focusing on the highest yielding crops, utilising various

Grenze ID: 01.GIJET.10.2.515_1

ways to assist solve agricultural challenges. For this classification, some algorithms have been used: decision tree, Naïve Bayes, logistic regression, random forest and Support vector machine (SVM). By comparing all these algorithms, we found that Naïve Bayes and random forest algorithm has a better accuracy rate as compared to others. In this project, a complete evaluation of the soil is being conducted based upon these respective factors. For crop forecast based on season, the crop selection methodology has been devised. In light of this, crops that are sown during the conclusion of the monsoon or at the start of the winter season—such as those planted between September and October—are referred to as Rabi and Kharif, respectively. These crops are referred to as monsoon crops. The seasonal crops will be forecast for Zaid, the short season that runs from March to July between the months of Rabi and Kharif. Rainfall is one of the primary elements influencing crop development. Other aspects that are considered include temperature, transpiration, land evaporation, and so on.

III. METHODOLOGY

Farmers first use a website or mobile application to access the platform and begin the login procedure. After their successful login, clients are presented with an extensive dashboard that is easily navigable and provides them with a variety of choices to solve their agricultural issues. These choices—soil analysis, crop forecasting, and fertiliser advice—all play crucial roles in raising agricultural output and efficiency.

The first step in the procedure for farmers interested in crop prediction is a laboratory investigation of the soil. This stage is crucial because it offers important information about the fertility and composition of the soil and acts as a basis for further investigations. A thorough report is produced after the soil study is finished, outlining the most important discoveries and observations. This report forms the foundation for additional study and the creation of recommendations.

An innovative algorithm is used to facilitate crop prediction and recommendation creation after the soil study. This method considers a number of variables, including as levels of nitrogen, phosphorus, and potassium, as well as environmental elements like humidity and rainfall. The system uses advanced data analytics techniques to produce customised crop suggestions based on the unique requirements and land characteristics of each farmer.

IV. ALGORITHMS

1) Random Forest: The Random Forest method is a supervised learning technique that enhances overall accuracy and manages overfitting by aggregating predictions from numerous decision trees.

A. Data Preparation

The training set is split up into many replacement subsamples. A random selection of characteristics is chosen for every subsample.

B. Decision Tree Construction

On each subsample, a decision tree is constructed using the chosen characteristics. The tree continues to grow until a certain depth or other condition is satisfied. The optimal split is selected at each node according to an impurity measure.

C. Prediction

For a new data point, each decision tree in the forest makes a prediction.

2)Naive Bayes algorithm:

D. Data Preparation

- Prepare your labelled dataset.

E. Probability Estimation

- Naive Bayes directly estimates the probabilities of each class and the conditional probabilities of features given each class from the training data, assuming feature independence.

F. Probability Calculation

- Calculate the prior probabilities of each class based on their frequencies in the training data.
- Calculate the likelihood of each feature value given each class.

G. Prediction

- When predicting the class for a new data point, calculate the posterior probability of each class given the features using Bayes' theorem.
- Choose the class with the highest posterior probability as the predicted class.

H. Evaluation

Evaluate the model's performance using appropriate metrics like accuracy, precision, recall, or F1 score.

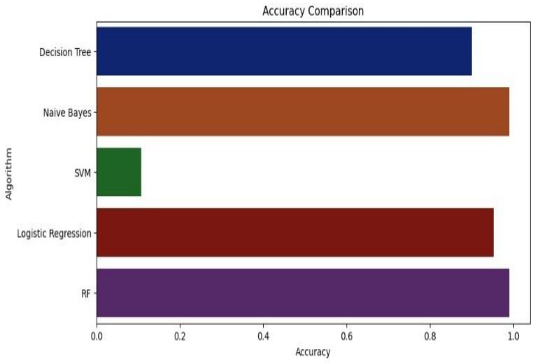

Figure. 1 Accuracy Comparison

```
accuracy_models = dict(zip(model, acc))
for k, v in accuracy_models.items():
    print (k, '-->', v)
```

```
Decision Tree --> 0.9
Naive Bayes --> 0.990909090909091
SVM --> 0.10681818181818181
Logistic Regression --> 0.9522727272727273
RF --> 0.990909090909091
```

Figure. 2 Accuracy Values

V. Datasets Description

The dataset utilized in this study encompasses environmental and soil factors crucial for crop cultivation, including temperature, rainfall, soil type, pH level, and nutrient content. It specifically focuses on three prominent crops: rice, maize, and watermelon. Each entry in the dataset contains feature values corresponding to these factors for each crop, along with class labels indicating the crop type. These factors play a pivotal role in determining the suitability of a particular crop for cultivation in a given region. This dataset can be used by researchers to train machine learning algorithms, like Naive Bayes, to forecast which crop would be best to grow given the current environmental and soil conditions.

	N	P	K	temperature	humidity	ph	rainfall	label
210	58	66	79	20.993736	19.334704	8.718193	93.552801	chickpea
211	43	66	79	19.462340	15.225390	7.976608	74.585651	chickpea
212	58	63	81	19.813445	14.697653	6.515500	78.965147	chickpea
213	23	62	85	18.974248	19.516122	8.490127	80.710875	chickpea
214	27	62	77	18.197370	14.710705	6.576416	70.181852	chickpea
215	28	72	84	18.729631	19.181973	6.481783	71.580102	chickpea
216	50	56	76	20.995022	19.860130	7.966605	73.507340	chickpea
217	39	71	84	20.281559	16.395352	8.140825	82.523397	chickpea
218	25	78	76	17.480426	15.755940	7.228963	66.969806	chickpea
219	31	70	77	20.888187	14.323138	6.492546	90.462283	chickpea

Figure. 3 Chickpea Dataset

	N	P	K	temperature	humidity	ph	rainfall	label
1302	105	30	50	25.299547	81.775276	6.376201	57.041471	watermelon
1303	114	8	50	24.746313	88.308663	6.581588	57.958261	watermelon
1304	93	22	52	26.587407	81.325632	6.932740	41.875400	watermelon
1305	80	26	55	24.534426	88.989272	6.140099	49.116187	watermelon
1306	85	27	45	26.071376	88.728566	6.467096	57.796528	watermelon
1307	85	22	53	25.965342	89.770767	6.849472	59.463386	watermelon
1308	82	22	45	26.223380	85.348660	6.512196	54.601593	watermelon
1309	118	13	54	24.413119	89.815740	6.039585	44.078435	watermelon
1310	83	25	53	26.491953	80.046782	6.057697	57.727992	watermelon
1311	86	15	47	24.043558	84.184068	6.423899	53.789300	watermelon

Figure. 4 Watermelon Dataset

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
5	69	37	42	23.058049	83.370118	7.073454	251.055000	rice
6	69	55	38	22.708838	82.639414	5.700806	271.324860	rice
7	94	53	40	20.277744	82.894086	5.718627	241.974195	rice
8	89	54	38	24.515881	83.535216	6.685346	230.446236	rice
9	68	58	38	23.223974	83.033227	6.336254	221.209196	rice

Figure. 5 Rice Dataset

VI. RESULT

Five machine learning techniques for crop prediction were compared in this study: Naive Bayes, Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM). The dataset encompassed crucial environmental parameters such as rainfall, humidity, soil nutrients, and others, with a focus on three main crops: watermelon, rice, and chickpeas. Our experiments revealed that Naive Bayes and Random Forest consistently outperformed the other algorithms in terms of prediction accuracy. Both Naive Bayes and Random Forest achieved the highest accuracy rates among the tested algorithms, demonstrating their effectiveness in crop prediction tasks.

The visualization of crop prediction outputs, as shown in the image below, illustrates the practical applicability of the developed models. The output depicts the predicted crops based on the given environmental parameters, providing farmers with valuable guidance for optimal crop selection and cultivation practices.

```
data = np.array([[23,72, 80, 19.02061277, 17.13159126, 6.920251378, 79.92698081]])
prediction = RF.predict(data)
print(prediction)

['chickpea']
```

Figure. 6 Chickpea Prediction

```
data = np.array([[85,27, 45, 26.87974371, 88.13159126, 6.520251378, 57.92698081]])
prediction = RF.predict(data)
print(prediction)

['watermelon']
```

Figure. 7 Watermelon Prediction

```
data = np.array([[90,42, 43, 20.87974371, 82.13159126, 6.520251378, 202.92698081]])
prediction = RF.predict(data)
print(prediction)

['rice']
```

Figure. 8 Rice Prediction

VII. CONCLUSIONS

The research presented in this paper highlights the critical role of machine learning algorithms in the domain of crop prediction. Through a meticulous comparative analysis, we underscored the importance of selecting the most appropriate algorithms tailored to the task at hand. By considering a diverse range of environmental parameters and evaluating multiple algorithms, we provided compelling evidence of the efficacy of Naive Bayes and Random Forest in accurately predicting the suitability of watermelon, rice, and chickpeas for cultivation.

Our findings carry substantial implications for agricultural decision-making practices. By furnishing farmers with valuable insights into crop selection strategies, we empower them to make informed decisions aimed at optimizing yields and efficiently utilizing resources. This knowledge enables farmers to adapt their cultivation practices to suit prevailing environmental conditions, thereby mitigating risks associated with crop failure and enhancing overall productivity.

REFERENCES

[1] Prabhu, S., Revandekar, P., Shirdhankar, S., & Paygude, S. (n.d.). Soil Analysis and Crop Prediction. Department of Electronics Engineering, KJ Somaiya Institute of Engineering and Information Technology, Sion, Maharashtra, India.

[2] T. T., Lokannavar, S., J., V., K. R., S., & Malipatil, S. B. (n.d.). Soil Quality Analysis and Crop Recommendation. Computer Science of Engineering, AMC Engineering College, Bengaluru, India.

[3] Varsha, V. M. Midhuna, Rilty, and R. Divya, "Agrotech: Soil Classification and Crop Recommendation," Student, Department of Computer Science and Engineering, Sahrdaya College of Engineering and Technology, Thrissur, India;

Assistant Professor, Department of Computer Science and Engineering, Sahrdaya College of Engineering and Technology, Thrissur, India, *Corresponding author: varshaathilak@gmail.com.

[4] K. Sunitha, G. Tarun Kumar, S. Saifulla, B. S. Sharath Kumar, and M. Bhat Venugopal, "Soil Analysis and Crop Recommendation using Machine Learning," Department of Information Science and Engineering, RNS Institute of Technology.

[5] Sadia Afrin, Abu TalhaKhan.“Analysis of Soil Properties and Climatic Data to Predict Crop Yields ” (2018 IEEE ICIS 2018).

[6] Amrutha A, Lekha R, A Sreedevi “Automatic Soil Nutrients Detection and Fertilizers Dispensary System” 2016 IEEE International Conference on Robotics: Current Trends and Future Challenges (RCTFC)