International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Enhancing Agricultural Efficiency in Tamil Nadu: A Machine Learning Framework for Crop and Irrigation Type Selection

*Thanishka. S[1], Dr. Jerline Amutha. A[2]**

[1] Student, PG Department of Computer Science and Technology, Women's Christian College, Chennai, Tamil Nadu, India

[2] Associate Professor, PG Department of Computer Science and Technology, Women's Christian College, Chennai, Tamil Nadu, India

ABSTRACT :

Though agriculture remains as an important sector in Tamil Nadu's economy, diversified environmental variables such as temperature, humidity, soil type, and climate make varied complications in the decisions of farmers regarding apt crops and irrigation types. The mentioned problem cannot be solved using traditional tools or decision-making. In this paper, a data-driven framework through machine learning models for the prediction of the right crops to be grown along with the suitable irrigation types is established based on district-specific environmental data. All the datasets were fetched from Tamil Nadu Agricultural University (TNAU) and filtered for missing values, label encoded, and normalized. The same datasets were trained to predict with multiple machine learning algorithms like Random Forest, SVC, and Gradient Boosting. The developed models are hyperparameter-tuned for better accuracy. In doing so, this interactive computational environment seeks to empower farmers with insight for better decisions towards increases in agricultural productivity and sustainability.

Keywords: Agriculture, Tamil Nadu's economy, Temperature, Climate, Crops, Irrigation Type

Introduction :

Agriculture forms part of the top two or three economic sectors in the state Tamil Nadu. A large section of people relies on agriculture as a source of living. However, crop cultivation and irrigation practices would be different district-wise in the state due to dissimilar climatic conditions, soil types, and availability of water supply. Tamil Nadu has nearly 36 major crops-paddy, sugarcane, cotton, maize, and many pulses-each under a specific set of environmental and soil conditions. Traditional decision making in agriculture often is too crude to allow it to deal with the complexities introduced by such diverse agricultural ecosystems.

Severe consequences of the wrong crop selection are those that farmers face. For instance, in such a case, where a crop is inappropriate to the type of soil, climate or irrigation source of a district yields will be low, higher input costs, and the risk of crop failure is always present; this will consequently lead to massive losses in finances. It is also the poor use of resources that may aggravate the problem of water shortage or wastes fertilizers hence further worsening productivity. It thus becomes relevant to take insights from farmers based on data specific for each district.

This paper, as such, presents a machine learning approach for the solution of these problems by predicting the best crop and irrigation practices for districts of Tamil Nadu. The analysis is undertaken by models such as Random Forest, SVC, and Gradient Boosting on various features like soil composition, climatic conditions, and previous crop performance. Application of these models and models will give the farmers the reliable tool for making decisions, which could recommend which crop to use and which suitable irrigation method must be employed at their specific location, thereby improving productivity in agriculture with ensured resource sustainability. Risks from the wrong crop selection are also reduced, and the general efficiency of farming practices in Tamil Nadu increases.

Literature Review :

The research paper **(Snehal S. Dahikar, 2014**) is novel for smart agriculture because it integrates weather prediction and crop selection into their model of crop prediction. Improved accuracy is achieved using an LSTM RNN in weather analysis compared with ANN, with RMSE values 5.023% for minimum temperature, 7.28% for maximum temperature, and 8.24% for rainfall. For crop selection, the accuracy of a Random Forest Classifier is impressive, with an accuracy of 97.235%, which makes the model adaptable in any geographical area with scope for further improvement. Similarly, study paper **(Elbasi, 2023)** explains the efficiencies of various algorithms that may be used in classification, including Naïve Bayes, Random Forest, and Multilayer Neural Networks, for the prediction of crops using various data collected from the different farms. This paper says the IoT devices that are needed in making agricultural operations optimize properly, with accurate crop classification through real-time data.

The Research paper **(P, 2021)** use an SVM model for crop and fertilizer prediction. It underlines a notable strength in separating data into classes by the construction of hyperplanes in high-dimensional space. The SVM model includes pivotal elements such as temperature, humidity, pH of soil, as well as the predicted amount of rainfall for each respective field of the farm to provide exact suggestions about the choice of crops and fertilizers.]. Lastly,

(Rajpoot, 2024)offers a holistic ML and IoT-driven model that suggests the appropriate crop, fertilizer, and irrigation system that would optimize agricultural productivity. Among the classifiers tested, the most accurate is the K-NN and further confirms the role of machine learning in agriculture.
Critical factors that affect crop growth were identified and integrated into the prediction model, based on the summary of the general outcome of findings indicated in (**Bochtis, 2021**) . In this research, core characteristics include the essential elements of nutrients in the soil, namely Nitrogen, Phosphorus, and Potassium, climatic factors in terms of temperature, humidity, and rainfall, and pH levels of the soil, which are known to define crop yields. By including these parameters in our model, we got an improved model which better predicts crop outputs under different environmental setups. The work by (**Paolini**) forms the strong basis of choosing such variables that play significant roles in crop growth in the model and keeping the suggested outcomes valid by the agricultural science.
Such a comprehensive review of previous literature helped provide the necessary guidelines for classifying and predicting the type of irrigation in this particular study. The derivation of the different irrigation methods, their classification, and possibly their use in various types of agricultural setting came from the work quoted in **(Bellvert, 2019)**. This is how, through this source alone, important information regarding the existence of different types of irrigation systems, specifically surface, drip, and sprinkler irrigation systems, came to the fore and was essential for developing the prediction model. The understanding of kinds of irrigation from the previous set of data improved upon getting the knowledge from **(Bellvert, 2019).** This improved the accuracy of the prediction of types of irrigation based on key environmental and soil parameters, thus improving the understanding of the irrigation practice.
Thus, in conclusion, I would shift to the implementation of models such as Random Forest, SVM, and Gradient Boosting with SMOTE to handle class imbalance as they have been observed to perform very well on crops and irrigation prediction.

Methodology :

3.1 Data Collection

For this research, data were collected specifically for the state of Tamil Nadu, which is known for its diverse agricultural practices and varied environmental characteristics. The dataset includes numerous columns representing critical agricultural and environmental variables. These columns are: **State Name**, **District Name**, **Crop**, **Area**, **Production**, **Maturity Period (Days)**, **Germination Period (Days)**, **Flowering Period (Days)**, **Vegetative Period (Days)**, **Season**, **Irrigation Type**, **Water Requirement (mm per season)**, **Frequency of Irrigation**, **Soil Type (Days)**, **Temperature (°C)**, **Humidity (%)**, **Soil pH**, **Nitrogen (kg/ha)**, **Phosphorus (kg/ha)**, **Potassium (kg/ha)**, **Climate**, and **Season**. The dataset was sourced from the Tamil Nadu Agricultural University (TNAU), a trusted institution that provides valuable agricultural data, ensuring the reliability and relevance of the information used in this analysis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	State_Nan	District_Na	Season	Crop	Area	Productior	Irrigation T	Water Req	Germinatic	Vegetative	Flowering	Maturity P	Soil Type	Frequency	Temperatu	Humidity ('	Soil pH	Nitrogen (l	Phosphoru	Potassium	Climate
2	tamil nadu	ariyalur	whole yea	bajra	565	1011345	Sprinkler Ir	400-600	25	50	75	100	Loamy	5	28	60	6.5	50	40	60	Tropical
3	tamil nadu	coimbator	whole yea	bajra	518	910	Sprinkler Ir	400-601	25	50	75	101	Loamy	5	28	60	6.5	50	40	60	Tropical
4	tamil nadu	cuddalore	whole yea	bajra	6184	15120	Sprinkler Ir	400-602	25	50	75	102	Loamy	5	28	60	6.5	50	40	60	Tropical
5	tamil nadu	dharmapu	whole yea	bajra	2669	4600	Sprinkler Ir	400-603	25	50	75	103	Loamy	5	28	60	6.5	50	40	60	Tropical
6	tamil nadu	dindigul	whole yea	bajra	7888	15870	Sprinkler Ir	400-604	25	50	75	104	Loamy	5	28	60	6.5	50	40	60	Tropical
7	tamil nadu	erode	whole yea	bajra	1986	2890	Sprinkler Ir	400-605	25	50	75	105	Loamy	5	28	60	6.5	50	40	60	Tropical
8	tamil nadu	kanchipura	kharif	bajra	12	34	Sprinkler Ir	400-606	25	50	75	106	Loamy	5	28	60	6.5	50	40	60	Tropical
9	tamil nadu	karur	whole yea	bajra	18003	5480	Sprinkler Ir	400-607	25	50	75	107	Loamy	5	28	60	6.5	50	40	60	Tropical
10	tamil nadu	krishnagiri	kharif	bajra	186	260	Sprinkler Ir	400-608	25	50	75	108	Loamy	5	28	60	6.5	50	40	60	Tropical
11	tamil nadu	madurai	whole yea	bajra	1020	1290	Sprinkler Ir	400-609	25	50	75	109	Loamy	5	28	60	6.5	50	40	60	Tropical
12	tamil nadu	nagapattin	kharif	bajra	3	1	Sprinkler Ir	400-610	25	50	75	110	Loamy	5	28	60	6.5	50	40	60	Tropical
13	tamil nadu	namakkal	whole yea	bajra	1512	2240	Sprinkler Ir	400-611	25	50	75	111	Loamy	5	28	60	6.5	50	40	60	Tropical
14	tamil nadu	perambalu	whole yea	bajra	4127	4310	Sprinkler Ir	400-612	25	50	75	112	Loamy	5	28	60	6.5	50	40	60	Tropical
15	tamil nadu	pudukkott	whole yea	bajra	16	40	Sprinkler Ir	400-613	25	50	75	113	Loamy	5	28	60	6.5	50	40	60	Tropical
16	tamil nadu	ramanatha	whole yea	bajra	720	800	Sprinkler Ir	400-614	25	50	75	114	Loamy	5	28	60	6.5	50	40	60	Tropical
17	tamil nadu	salem	whole yea	bajra	5957	12250	Sprinkler Ir	400-615	25	50	75	115	Loamy	5	28	60	6.5	50	40	60	Tropical
18	tamil nadu	sivaganga	whole yea	bajra	64	70	Sprinkler Ir	400-616	25	50	75	116	Loamy	5	28	60	6.5	50	40	60	Tropical
19	tamil nadu	thanjavur	kharif	bajra	3	7	Sprinkler Ir	400-617	25	50	75	117	Loamy	5	28	60	6.5	50	40	60	Tropical

Figure 3.1: Illustration of few rows of collected dataset

3.2 Data Preprocessing

3.2.1 Handling Missing Values

Imputation methods were utilized for missing data to protect the dataset from incomplete data. As for the numerical columns, the mean imputation was used, where the missing values are filled with average values of the available data. In case of a categorical column, the missing values were filled with the most frequent category in each respective column. This method in turn protected the data structure and the data properties of the dataset.

3.2.2 Encoding Categorical Variables

The Label Encoder was the technique used to transform categorical variables to numerical ones. The label encoder method was used to transfer the unique names into the numerical format that is necessary for modelling. All the remaining categorical variables such as the district names, soil types, climates, seasons, crops, and irrigation methods were converted into numerical labels as a way of acidifying their easy usage in the modelling process.

3.2.3 Scaling Numerical Features

The numerical columns underwent standardization with the Standard Scaler function. This ensures that the mean of the data equals 0 and the standard deviation is 1. Thus, the models using the Euclidean distance are the ones that benefit from this procedure as it puts all the features on an equal-footing scale.

3.2.4 Addressing Class Imbalance

Proposed particularly in the Irrigation Type forecast, SMOTE (Synthetic Minority Over-sampling Technique) was used to handle class imbalance in the dataset. The SMOTE algorithm was introduced to the data that included both majority and minority classes, and the data were balanced before the model was built. This method not only increased the model's accuracy in predicting the minority class but it also helped to reduce bias towards the majority class and thus provided a better prediction in general.

3.2.5 Column Transformer for Simultaneous Preprocessing

The Column Transformer was employed to run the necessary numeric and categorical preprocessing steps simultaneously. This was achieved by the combination of various procedures into one process, thus facilitating more fine-grained and convenient data preparation, and finally all the required preprocessing steps were carried out in a single step.

3.2.6 Preprocessing Pipeline for Consistent Transformation

A preprocessing pipeline was established to make the entire process of data transformation automatic. The data missing values imputation, encoding, scaling, and class balancing were applied in the same way to both the training and test datasets, ensuring consistency and reliability throughout the model training process.

3.3 Feature Engineering

The aim of feature selection was to enhance the prediction model with machine learning algorithms, by highlighting the key features that were used to forecast the Crop and Irrigation Type. The adopted methodology that was used for the analysis was Recursive Feature Elimination (RFE) with a RandomForestClassifier as the base estimator. The process was that the algorithm ranked and removed the least important features based on a 0 - 1 ranking with separate variables.

Temperature (° C), Humidity (%), Soil pH, Nitrogen (kg/ha), Phosphorus (kg/ha), Potassium (kg/ha), Soil Type, Climate and Season are the most eminent features derived from the process. Thus, they were chosen as the major factors that decide the model's performance and they were preserved for an additional test. The categorical features (Soil Type, Climate, and Season) were Label Encoder encoded to convert them into a numerical form that is appropriate for machine learning models.

Features that are mostly not significant for the prediction of models like Maturity Period (Days), Germination Period (Days), Flowering Period (Days), Vegetative Period (Days), Area, Production, Water Requirement (mm per season), and Frequency of Irrigation, were removed. Furthermore, it was through this process that we could make our data more available to speed up the machine learning process and diminish the highly accurate prediction.

	A	B	C	D	E	F	G	H	I	J	K
1	District_N	Season	Crop	Soil Type	Temperatu	Humidity (	Soil pH	Nitrogen (	Phosphoru	Potassium	Climate
2	ariyalur	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
3	coimbator	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
4	cuddalore	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
5	dharmapu	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
6	dindigul	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
7	erode	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
8	kanchipura	Kharif	bajra	Loamy	28	60	6.5	50	40	60	Tropical
9	karur	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
10	krishnagiri	Kharif	bajra	Loamy	28	60	6.5	50	40	60	Tropical
11	madurai	Whole Yea	bajra	Loamy	28	60	6.5	50	40	60	Tropical
12	nagapattin	Kharif	bajra	Loamy	28	60	6.5	50	40	60	Tropical

Figure 3.2 Illustration of the Pre-Processed Dataset After Feature Engineering

3.4 Dataset Splitting

To analyse the agricultural pattern in Tamil Nadu, the dataset was structured in such a way that it could be split into two parts. The crop dataset had some crucial features such as District Name, Crop, Temperature (°C), Humidity (%), Soil pH, Nitrogen (kg/ha), Phosphorus (kg/ha), Potassium (kg/ha), Climate, Season, and soil type. This large set of parameters helps in the analysis of reasons for the crop choices by districts. Meanwhile, the irrigation set was collated to just the contents of the Crop and Irrigation Type columns only, thus simply giving an evident presentation of how the crop choices are being associated with the irrigation schemes used. This systematic manner of dealing with the problem not only reinforces its clarity but also improves the accuracy of the predictive model on crop and irrigation patterns in Tamil Nadu.

3.5 MACHINE LEARNING ALGORITHMS IN CROP AND IRRIGATION PREDICTION

The experiments employed models like Random Forest Classifier, Support Vector Classifier (SVC), and Gradient Boosting to analyse the two datasets, among others: crop prediction and irrigation type classification. These models were tested under the effect of Synthetic Minority Over-Sampling Technique (SMOTE) to deal with the class imbalance both with and without it. Among others, SMOTE was a particularly good approach to tackling the imbalanced data distribution and making sure that the models received training data that was balanced and gave them the ability to generalize.

3.5.1 Random Forest Classifier

The Random Forest model, which includes an ensemble of multiple decision trees, performed well in detecting complex relationships in the data. Ensemble learning methods build more robust models that account for variance in the dataset, particularly important with noisy datasets or datasets with many variables, such as the crop and irrigation datasets used in this research.

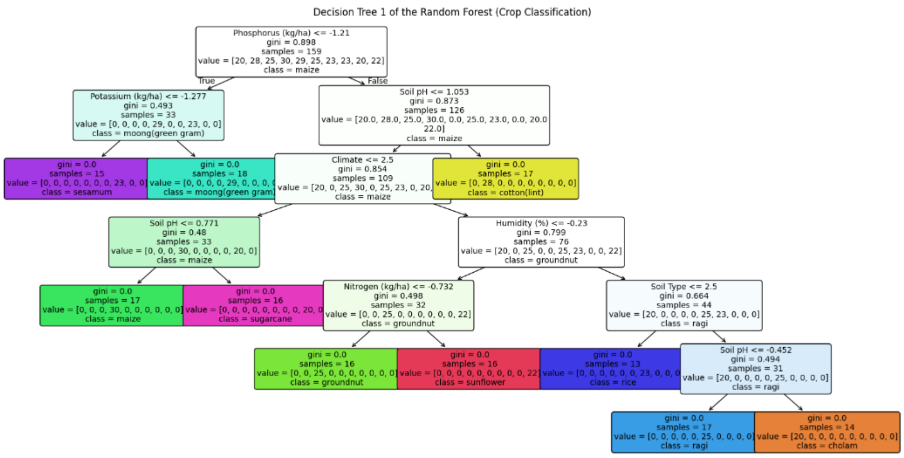


Figure 3.3: Decision Tree model for Crop Classification

3.5.2 Support Vector Classifier (SVC)

The Support Vector Classifier (SVC) was also applied to the analysis, which separates classes by maximizing the margin between classes. SVC is effective when classes are not linearly separable, particularly with a kernel function, making it well suited for classifying crops and irrigation.

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1, \forall i$$

Where w is the normal vector to the hyperplane, b is the bias, and yi represents the labels.

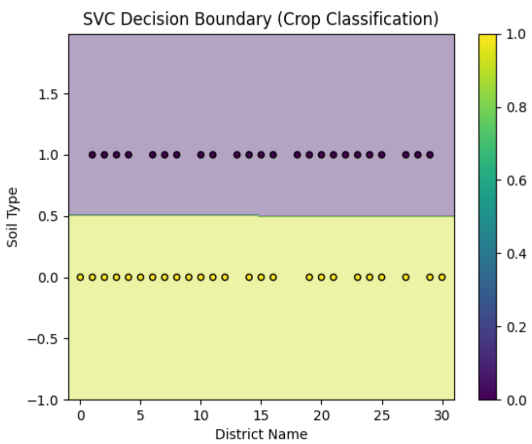


Figure 3.4: Support Vector Classifier model for Crop Classification

3.5.3 Gradient Boosting

Gradient Boosting was implemented, such that the model learns iteratively by adding weak learners together to reduce residuals. The algorithm allows the model to reduce prediction error incrementally as it learns from each iteration.

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

where Fm(x) is the updated model, Fm−1(x) is the previous iteration of the model, γ is the learning rate, and hm(x) is the weak learner added at each step

3.6 Model Training

The experiments employed models like Random Forest Classifier, Support Vector Classifier (SVC), and Gradient Boosting to analyse the two datasets, among others: crop prediction and irrigation type classification. These models were tested under the effect of Synthetic Minority Over-Sampling Technique (SMOTE) to deal with the class imbalance both with and without it. Among others, SMOTE was a particularly good approach to tackling the imbalanced data distribution and making sure that the models received training data that was balanced and gave them the ability to generalize.

3.7 Model Performance Evaluation

Each model's performance was compared with and without SMOTE, an oversampling application that improved the model's ability to handle following class imbalance for the crop and irrigation datasets. The evaluation included the following distinct metrics:

3.7.1 Accuracy

Accuracy is calculated as the proportion of all true positives (TP) and true negatives (TN) predictions made correctly by the model. Accuracy is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Where:
TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

3.7.2 Precision

Precision calculates the accuracy of the positive prediction.

3.7.3 Recall (Sensitivity)

Recall is the metric that estimates the model's California acquisition of identifying all relevant flows.
as a metric simply states how many of the actual positives were correctly identified by the TP predictions.

3.7.4 F1 Score

The F1 Score is the average between precision and recall. This metric is useful for class imbalance datasets.

3.7.5 Support

Support is the number of actual occurrences of each class in the dataset. Support provides context to the precision and recall metrics to expose how many instances exist for each class. In other words, support is defined as:
Support = TP + FN
The computation of support indicates how many true instances exist for each class, therefore providing context to interpret the model's performance with other metrics.

3.8 Implementation

The experiments employed models like Random Forest Classifier, Support Vector Classifier (SVC), and Gradient Boosting to analyse the two datasets, among others: crop prediction and irrigation type classification. These models were tested under the effect of Synthetic Minority Over-Sampling Technique (SMOTE) to deal with the class imbalance both with and without it. Among others, SMOTE was a particularly good approach to tackling the imbalanced data distribution and making sure that the models received training data that was balanced and gave them the ability to generalize.

Results And Discussion

For crop prediction, the accuracy of model Random Forest on the original dataset was only 86.21%, which was pretty good but had problems with underrepresented classes. When SMOTE was applied to re-balance the dataset, the accuracy rose to 95.86%, revealing that the model, henceforth, handled imbalanced data much better and made more reliable, balanced predictions over all the crops. Gradient Boosting performed well with accuracy equal to 94.48% on the original data and 95.17% with SMOTE applied. The SVC model shows the least performance of the three models with an accuracy of 82.76% on the original data but increased up to 92.41% with SMOTE application.

For irrigation type prediction, again, Random Forest gave the best model, the accuracy of 97.24% on the original dataset, and with SMOTE, its accuracy further increased to a surprising 99.31%, which is a sign of strength and effectiveness in predicting types of irrigation even with class imbalance. Gradient Boosting also gave decent accuracy by reaching 91.72% on the original data. This further increased with the usage of SMOTE to 95.17%. However, SVC still failed with only 53.79% accuracy on the original data, though it improved moderately to 57.24% with SMOTE.

4.1 Classification Report for Crop

Table 4.1: Classification Report for Random Forest (Original Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	0.08	1.00	0.15	1
2	0.00	0.00	0.00	2
3	0.00	0.00	0.00	2
5	1.00	1.00	1.00	3
6	1.00	1.00	1.00	3
...	...	...	...	...
Total	**Accuracy:**		0.8621	145

Table 4.2: Classification Report for Random Forest (SMOTE Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
2	0.50	0.50	0.50	2
...	...	...	...	...
Total	**Accuracy:**		0.9586	145

Table 4.3: Classification Report for Gradient Boosting (Original Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
2	0.00	0.00	0.00	2
3	1.00	1.00	1.00	2
5	1.00	1.00	1.00	3
6	1.00	1.00	1.00	3
...	...	...	...	...
Total	**Accuracy:**		0.9448	145

Table 4.4: Classification Report for Gradient Boosting (SMOTE Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
2	0.50	0.50	0.50	2
3	1.00	1.00	1.00	2
5	1.00	1.00	1.00	3
6	1.00	1.00	1.00	3
...	...	...	...	...
Total	**Accuracy:**		0.9517	145

Table 4.5: Classification Report for SVC (Original Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
2	0.00	0.00	0.00	2
3	0.50	0.50	0.50	2
...	...	...	...	...
Total	**Accuracy:**		0.8276	145

Table 4.6: Classification Report for SVC (SMOTE Data) on Predicting Crop

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
2	1.00	0.50	0.67	2
...	...	...	...	...
Total	**Accuracy:**		0.9241	145

4.3 Classification Report for Irrigation Type

Table 4.7: Classification Report for Random Forest (Original Data) on Predicting Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.80	1.00	0.89	4
2	1.00	1.00	1.00	4
3	1.00	1.00	1.00	35
4	0.84	1.00	0.91	16
5	1.00	0.89	0.94	38
6	1.00	1.00	1.00	3
7	1.00	1.00	1.00	44
Total	Accuracy:		0.9724	145

Table 4.8: Classification Report for Random Forest (SMOTE Data) Predicting on Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.80	1.00	0.89	4
2	1.00	1.00	1.00	4
3	1.00	1.00	1.00	35
4	1.00	1.00	1.00	16
5	1.00	0.97	0.99	38
6	1.00	1.00	1.00	3
7	1.00	1.00	1.00	44
Total	Accuracy:		0.9931	145

Table 4.9: Classification Report for SVC (Original Data) Predicting on Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	1
1	0.00	0.00	0.00	4
2	0.00	0.00	0.00	4
3	0.41	0.97	0.58	35
4	0.54	0.44	0.48	16
5	0.76	0.84	0.80	38
6	0.00	0.00	0.00	3
7	0.71	0.11	0.20	44
Total	Accuracy:		0.5379	145

Table 4.10: Classification Report for SVC (SMOTE Data) on Predicting Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	0.17	1.00	0.29	1
1	0.10	0.25	0.14	4
2	0.36	1.00	0.53	4
3	0.48	0.71	0.57	35
4	0.58	0.88	0.70	16
5	0.94	0.79	0.86	38
6	0.60	1.00	0.75	3
7	1.00	0.11	0.20	44
Total	**Accuracy:**		0.5724	145

Table 4.11: Classification Report for Gradient Boosting (Original Data) on Predicting Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.67	1.00	0.80	4
2	1.00	1.00	1.00	4
3	0.94	0.97	0.96	35
4	0.84	1.00	0.91	16
5	0.89	0.87	0.88	38
6	1.00	1.00	1.00	3
7	0.97	0.86	0.92	44
Total	**Accuracy:**		0.9172	145

Table 4.12: Classification Report for Gradient Boosting (SMOTE Data) on Predicting Irrigation Type

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.80	1.00	0.89	4
2	1.00	1.00	1.00	4
3	1.00	0.97	0.99	35
4	0.76	1.00	0.86	16
5	1.00	0.89	0.94	38
6	1.00	1.00	1.00	3
7	0.98	0.95	0.97	44
Total	**Accuracy:**		0.9517	145

4.2 Model Comparison

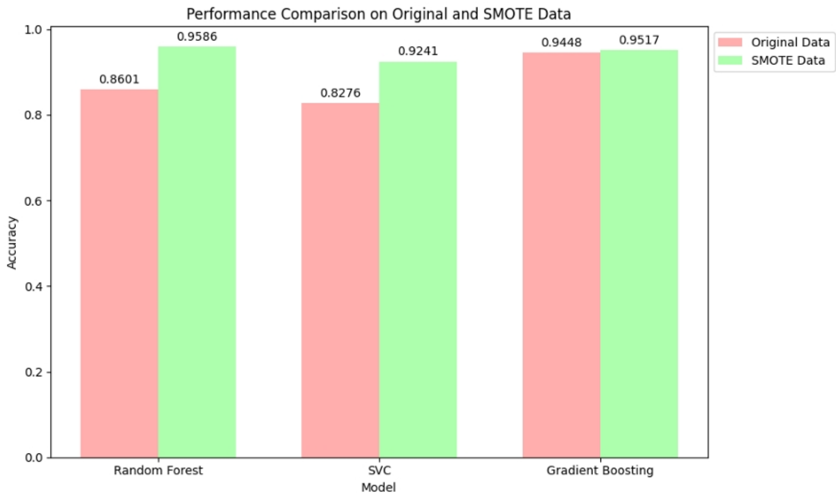


Figure 4.1: Model Comparison Plot for Crop Prediction.

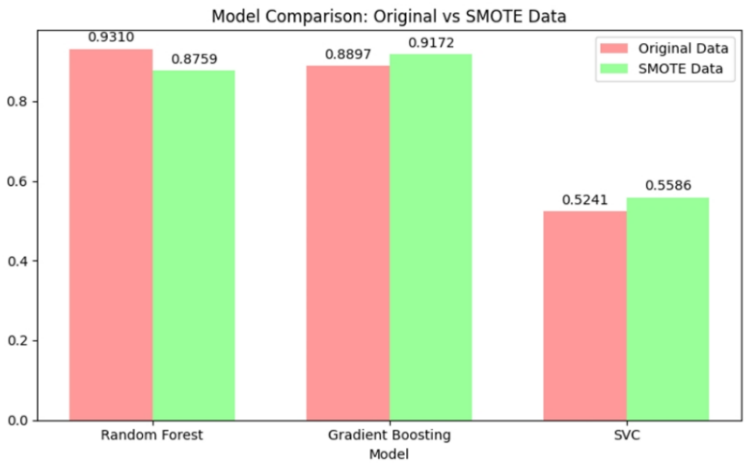
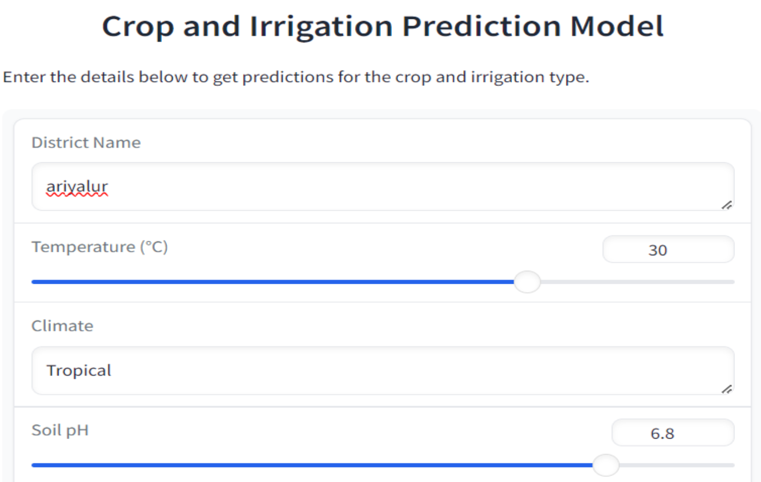


Figure 4.2: Model Comparison Plot for Irrigation Type.

4.3 User Interface Design



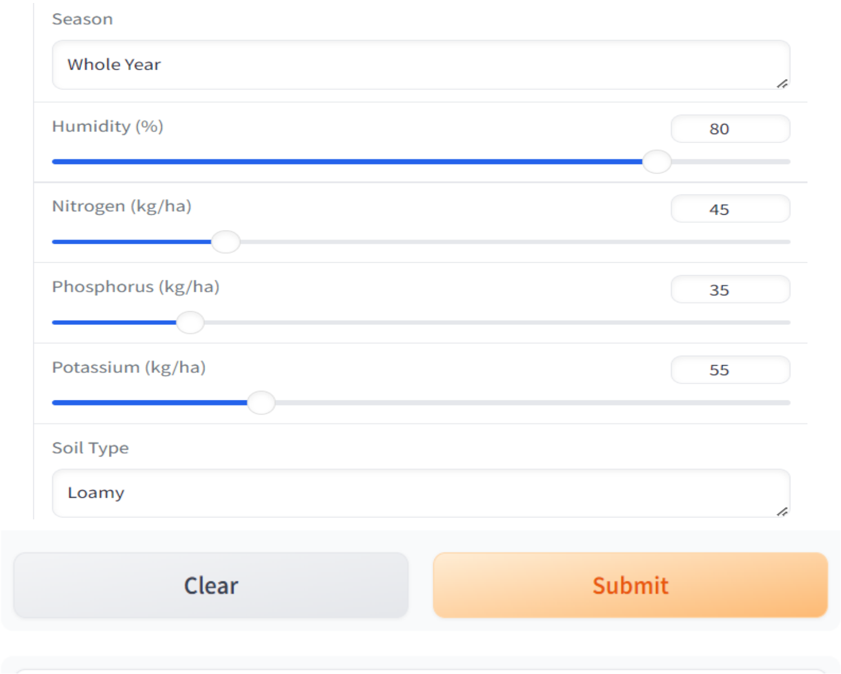


Figure 4.3: Gradio User Interface

4.4 Outcome

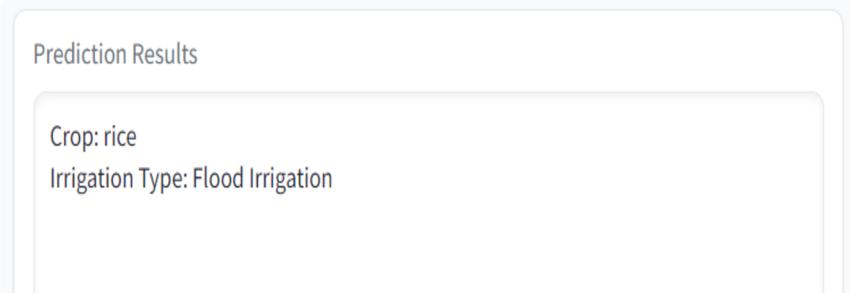


Figure 4.4: Gradio User Interface

Crop prediction is made using features like District Name, Temperature, Humidity, Soil Type, Nitrogen, Phosphorus, Potassium, Season, Climate, and Soil ph. The type of irrigation is then predicted based on the crop predicted.

CONCLUSION :

The best-performing model, concerning crop and irrigation-type predictions, was the Random Forest model with SMOTE, which attained 95.86% for crops and 99.31% for irrigation types. Therefore, it suggests that the appropriate use of SMOTE in conjunction with the Random Forest model can indeed tackle class imbalances effectively and enhance the predictability of the models or, in other words, output more balanced predictions. With a strong emphasis on precision, this model would be a safe choice both for crop-type prediction and for irrigation-type prediction, for which the future predictions will be accurate and generic enough to apply for all different distributions of the data.

FUTURE WORK :

In the future, the extension of this research would be done by exploring other machine learning models and advance techniques to ensure the improvement of precision and reliability in crop and type irrigation prediction. This could be done, for instance, with deep learning models like convolution neural networks and recurrent neural networks as a way of incorporating complex patterns from the data, especially with time-series climate variables like temperature, humidity, and rainfall.

Another avenue for future work includes enhancing the dataset scope through higher numbers of regions, soil types, and crops, which can add generality and applicability of the model in various farm environments. Feeding in data from IoT sensors in real-time can also lead to dynamic predictions that can be applied to the agriculture field for precision agriculture farming, where the farmers obtain timely information for proper decisions regarding crop selection, irrigating time, and fertilizer application. Such integration of sustainability metrics in the model, for example water us age efficiency, environmental impact, and cost-effectiveness, could aid development of more holistic decision-support systems promoting not only optimization of agricultural productivity but also eco-friendly as well as resource-efficient farming practices.

REFERENCES :

[1] Snehal S. Dahikar, Dr. Sandeep V. Rode, “Agricultural Crop Yield Prediction Using Artificial Neural Network Approach”, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2(1), 2014.

[2] Omolola M. Adisa, Joel O. Botai, Abiodun M. Adeola, Abubeker Hassen et al. “Application of Artificial Neural Network for Predicting Maize Production in South Africa”, Sustainability, 2019, 11(4).

[3] Niketa Gandhi, Owaiz Petkar, and Leisa J. Armstrong “Rice Crop Yield Prediction Using Artificial Neural Networks”, IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development, 2016.

[4] William W. Guo and Heru Xue, “Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models”, Mathematical Problems in Engineering, Hindawi, 2014. Suhas S. Athani and C.H. Tejeshwar, “Support Vector Machine-Based Classification Scheme of Maize Crop”, IEEE Conference, 2017.

[5] M. Safa, S. Samarasinghe, and M. Nejat, “Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New Zealand”, J. Agr. Sci. Technology, 2015, Vol. 17.

[6] Suhas S. Athani and C.H. Tejeshwar, “Support Vector Machine-Based Classification Scheme of Maize Crop”, IEEE Conference, 2017, 84-88. [16] Monali Paul, Santosh K. Vishwakarma, and Ashok Verma, “Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach”, IEEE International Conference on Computational Intelligence and Communication Networks, 2015.

[7] Monali Paul, Santosh K. Vishwakarma, and Ashok Verma, “Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach”, IEEE International Conference on Computational Intelligence and Communication Networks, 2015.

[8] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare, “Crop Recommendation System to Maximize Crop Yield Using Machine Learning Technique”, International Research Journal of Engineering and Technology, 2017, 4(12).

[9] S. Veenadhari, Bharat Misra, C.D. Singh, “Machine learning approach for forecasting crop yield based on climatic parameters”, International Conference on Computer Communication and Informatics, 2014, Coimbatore, India.

[10] Gonzalez-Sanchez, A., Frausto-Solis, J., Ojeda-Bustamante, W., “Attribute selection impact on linear and nonlinear regression models for crop yield prediction”, Sci. World Journal, 2014.

[11] M. A. Jayaram and Netra Marad, “Fuzzy Inference Systems for Crop Yield Prediction”, Journal of Int