IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Precision Farming Through Intelligent Crop And Fertilizer Prediction

[1]Sirugudu Rajasekhar, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions

[2]Dr. P. Sujatha, Associate Professor at Miracle Educational Society Group of Institutions

[3]Seera Sitalakshmi, Assistant Professor at Miracle Educational Society Group of Institutions

ABSTRACT

The accurate prediction of crop is crucial for effective agricultural planning and resource mobilization. This project constructs and enhances the system of predicting crop yield harnessing the features of the agriculture environment. The important environmental features to be focused on includes among other soil type, rainfall and temperature by implementing various feature selection methods including BORUTA and Recursive Feature Elimination (RFE). These features are fed to the ensemble of machine learning algorithms such as Random Forest, SVM and KNN to improve the prediction accuracy. The system also provides fertilizer application and yield prediction. Testing in databases yields promising results in the improvement of precision and decision making. The proposed model proves to be reliable and cost efficient and assists farmers with actionable information to enhance crop productivity and sustainable farming techniques.

Keywords: BORUTA, SVM, KNN

INTRODUCTION:

Precision agriculture remains a key aspect in terms of food and economic security. Developing precision agriculture through robust crop forecasting is an essential part of enhancing resource management, budgeting and total output. However, much attention has not yet been given to these traditional practices considering agricultural systems to be complex with respect to environmental factors like weeds, soil, climate and rainfall.

The work proposes an intelligent architecture model that applies machine learning algorithms to systems that are capable of making reliable agricultural forecasts. The model is provided the weather data, soil characteristics and historical farm yield. The model is able to calculate the crop type, amount of fertilizer required to achieve a particular crop yield. Further, it rectifies the class imbalance issues by oversampling techniques including SMOTE. This allows farmers to receive reliable suggesting services accordingly, cutting risk and increasing output. The project enhances agriculture by providing effective solutions to emerging issues with no geographical or economic limitations.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

Many experts and scientists have pointed out the lack of models needed for handling large datasets and dealing with various climatic conditions as the major reason for unfavourable results in forecasting. The applicability of various methods by other researchers highlights the issue of significant feature selection and classifier adjustment that allow damages caused by environmental factors to be predicted with greater accuracy optimal farming and planting dimensions.

Key Gaps:

1. Undesirable Feature Selection: Many of the models in Paprika have failed to recommend relevant environmental attributes that will affect the growth and yield of the crops.
2. Limited Classifier Performance: Availability of good algorithms has not been upheld in existing systems which leads to low accuracy and poor scaling of the models.
3. Dataset Imbalance: Class distribution imbalance in agricultural datasets has been proposed and practiced sometimes ignoring the class balance bias when making predictions.
4. Inability of recommending Fertilizers Incorporation: A small number of the systems have been developed that are able to provide fertilizer recommendations that suit soil and crops which limits the chances of making beneficial practices.
5. Limited intake and readjustment: Outdated models have a problem with operational and readjustment processes with newest information.

PROBLEM STATEMENT:

Preparing a set of crops models according to the environmental parameters is not straightforward as it involves a number of relevant features and complex models that need to be built.

Key Challenges:

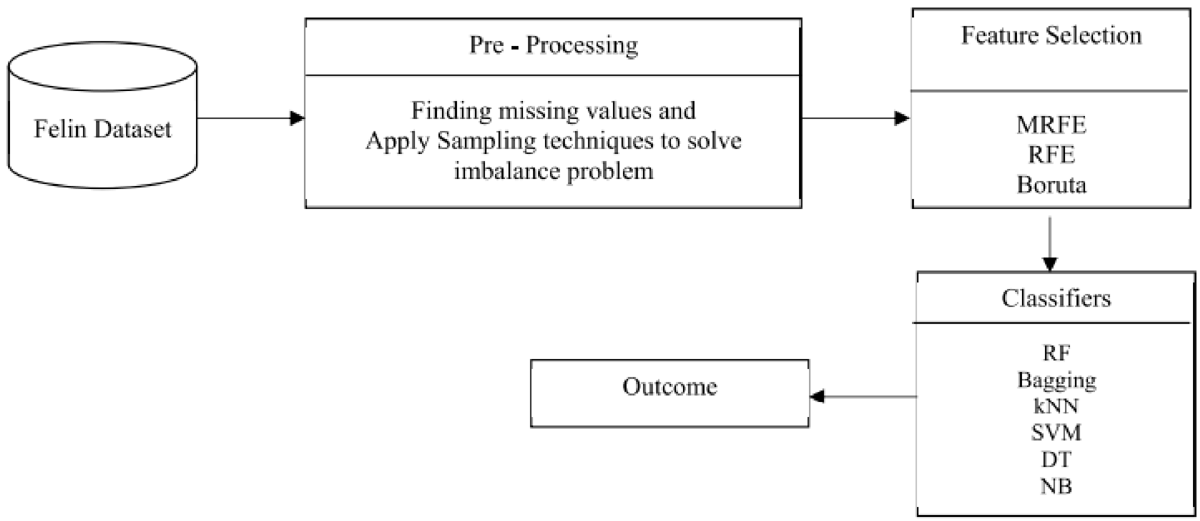
1. Feature Selection: Understanding and determining which environmental attribute is essential for the crop growth.
2. Model Accuracy: Making classifiers that will make the predictions true for a single or variety of datasets with the set of crops models.
3. Data Imbalance: Incorporating class imbalance problem contained in agricultural data which has an oversampling solution.
4. Fertilizer Recommendations Incorporation: Fulfilling the necessity of practical intelligence for herding plans.
5. Usefulness and adjustability: Nomenclature of the model precision to a dynamic atmosphere with real-time information.

PROPOSED METHOD:

The project suggests a machine learning enhanced artificial biological system with a proper feature selection and classification methods for predicting crops. The methodology includes:

1. **Data Preprocessing:** Data containing pricing or other environmental set meanings modification and standardization.
2. **Feature Selection:** Employing BORUTA and RFE algorithms to identify significant attributes.
3. **Classifier Training:** Implementing algorithms like Random Forest, SVM, and KNN to develop predictive models.
4. **Oversampling Techniques:** Addressing class imbalances with SMOTE and Random Oversampling.
5. **Fertilizer Recommendations:** Integrating KNN-based predictions for optimal fertilizer use.

ARCHITECTURE:



DATASET:

The dataset comprises environmental attributes such as soil type, rainfall, temperature, and humidity, alongside labeled crop data. It includes approximately 22 crop types with additional metadata for fertilizer recommendations. Preprocessing steps involve data normalization and shuffling to ensure uniformity. Class imbalances are addressed using techniques like SMOTE and Random Oversampling. Feature selection algorithms, including BORUTA and RFE, are applied to identify critical attributes influencing crop growth and yield.

METHODOLOGY:

• DB Pre-processing:

• The agricultural db that has environmental attributes and crop names and ids attached to them was imported.

• The data set was normalized and shuffled to avoid bias.

• The dataset was divided into training and testing set for the testing of the model.

• Feature Selection:

• Use BORUTA to ascertain the significant environmental features affecting crop yield.

• Employ RFE in the feature selection process in order to obtain the best possible model input attributes.

• Classifier Implementation:

• Multiple machine learning classifiers such as Random Forest, SVM, KNN, and Bagging Classifiers will be trained.

• Evaluate performance using precision, recall, F1-score and accuracy metrics.

• Oversampling Techniques:

• Counteract dataset imbalance using SMOTE and Random Oversampling.

• The classifiers will be retrained using oversampled datasets in order to increase performance.

• Fertilizer Prediction:

• KNN Model will be trained using the cleaned fertilizer text data.

• Based on the crop types predicted, suggest the fertilizer according to the soil and climatic conditions.

• Yield Prediction:

• Crops yield can be further estimated by extending the model using the particular environmental parameters selected.

• For accuracy check the predictions with the real data collected.

• Model Validation and Optimization:

• Evaluate how the classifier has performed over time based on various metrics.

• Improve hyperparameters to maximize accuracy as well as the scalability.

• System Deployment:

• Create a UI for farmers to enter data and get predictions from the system.

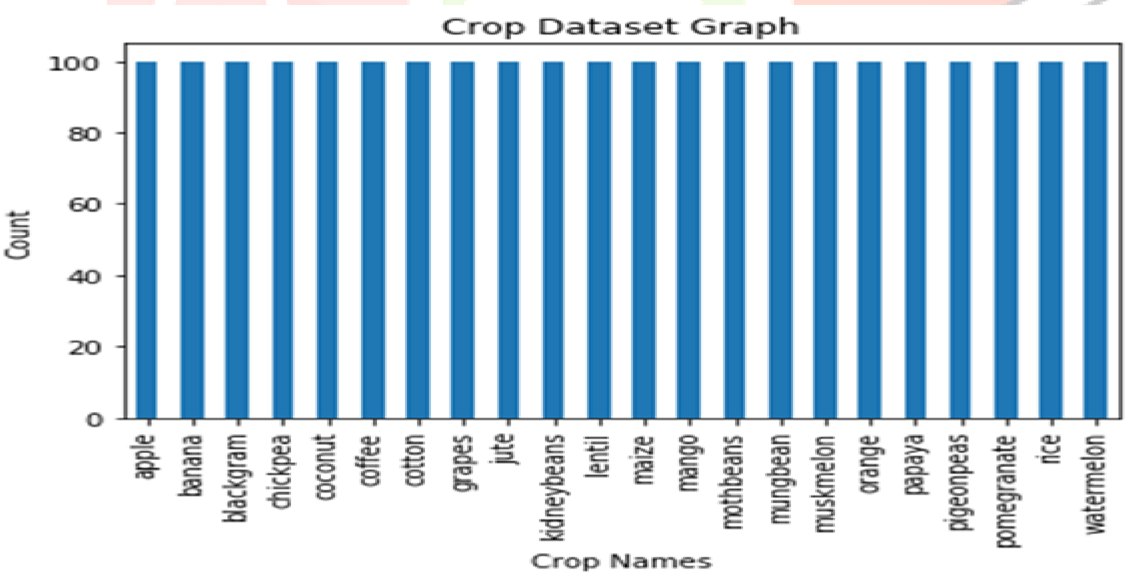
• The display allows for real time recommendations and prediction usage.

• Performance Analysis:

• Graphs and charts will be used to visualize the accuracy metrics.

• Showcase improvement by comparing the baseline models against the optimized models.

RESULTS:



Graph x-axis represents Crop Names in dataset and y-axis represents counts of records available for that crop

	Sampling Technique	Classifiers	Precison	Recall	FScore	Accuracy
0	Without Sampling	Naive Bayes	98.840326	98.856840	98.820364	98.863636
1	Without Sampling	Decision Tree	98.804441	98.901696	98.786229	98.863636
2	Without Sampling	SVM	98.840326	98.856840	98.820364	98.863636
3	Without Sampling	KNN	96.468574	96.467797	96.214773	96.363636
4	Without Sampling	Bagging Classifier	99.107471	99.140931	99.065819	99.090909
5	Without Sampling	Random Forest	99.282297	99.380165	99.277795	99.318182
6	Without Sampling	Feed Forward Neural Network	98.350901	98.333515	98.325077	98.409091

Train all algorithms without applying any SAMPLING or features selection algorithms and then we can see accuracy and other metrics for each algorithm

	Sampling Technique	Classifiers	Precison	Recall	FScore	Accuracy
0	SMOTE Sampling	Naive Bayes	99.179842	99.145827	99.133463	99.090909
1	SMOTE Sampling	Decision Tree	98.503788	98.386080	98.415365	98.409091
2	SMOTE Sampling	SVM	99.179842	99.145827	99.133463	99.090909
3	SMOTE Sampling	KNN	96.146947	96.092082	95.807161	95.909091
4	SMOTE Sampling	Bagging Classifier	99.207133	99.236271	99.206475	99.090909
5	SMOTE Sampling	Random Forest	99.368487	99.343455	99.339967	99.318182
6	SMOTE Sampling	Feed Forward Neural Network	98.737436	98.886621	98.760258	98.636364

Training all algorithms by applying SMOTE sampling on training features

	Sampling Technique	Classifiers	Precison	Recall	FScore	Accuracy
0	ROSE Sampling	Naive Bayes	98.868237	98.997551	98.905184	98.863636
1	ROSE Sampling	Decision Tree	99.493386	99.615200	99.542781	99.545455
2	ROSE Sampling	SVM	98.868237	98.997551	98.905184	98.863636
3	ROSE Sampling	KNN	97.590659	97.804315	97.588398	97.500000
4	ROSE Sampling	Bagging Classifier	99.059220	99.156063	99.094940	99.090909
5	ROSE Sampling	Random Forest	99.326599	99.372513	99.343546	99.318182
6	ROSE Sampling	Feed Forward Neural Network	99.245690	99.074074	99.140565	99.090909

Training all algorithms with Random Over Sampling

	Features Selection	Total Features	Selected Features	Classifiers	Precison	Recall	FScore	Accuracy
0	BORUTA	7	6	Naive Bayes	99.493386	99.493386	99.493386	99.545455
1	BORUTA	7	6	Decision Tree	98.681996	98.465473	98.505807	98.636364
2	BORUTA	7	6	SVM	99.493386	99.493386	99.493386	99.545455
3	BORUTA	7	6	KNN	98.564763	98.564763	98.505838	98.636364
4	BORUTA	7	6	Bagging Classifier	99.305081	99.254151	99.246344	99.318182
5	BORUTA	7	6	Random Forest	99.531025	99.493386	99.498674	99.545455
6	BORUTA	7	6	Feed Forward Neural Network	98.777416	98.789143	98.776831	98.863636

Accuracy and other metrics output obtained after applying BORUTA

	Features Selection	Total Features	Selected Features	Classifiers	Precison	Recall	FScore	Accuracy
0	RFE	7	5	Naive Bayes	99.494949	99.431818	99.429590	99.545455
1	RFE	7	5	Decision Tree	99.222978	99.331109	99.256555	99.318182
2	RFE	7	5	SVM	99.494949	99.431818	99.429590	99.545455
3	RFE	7	5	KNN	97.080751	96.709957	96.770479	97.045455
4	RFE	7	5	Bagging Classifier	99.783550	99.715909	99.742508	99.772727
5	RFE	7	5	Random Forest	100.000000	100.000000	100.000000	100.000000
6	RFE	7	5	Feed Forward Neural Network	98.623737	98.555195	98.523507	98.636364

Applying RFE features selection algorithm and then training all the algorithms and then Random Forest got 100% accuracy after applying RFE

```
SOIL Test Data : [ 99.          57.          35.          26.75754171  81.17734011
   5.96037006 272.2999056 ] PREDICTED CROP =====> rice
Predicted Yield = 248.35443606800033 Bags
rice fertilizers
```

We can see SOIL test data and then we can see predicted crop for test data is 'RICE'and in next line we can see predicted YIELDand in next line we can see required fertilizers

CONCLUSION

The outcome of this project is the creation of a smart crop prediction system. This system utilizes the use of feature selection methods and machine learning classifiers. Because of the appropriate techniques to tackle challenges like imbalances in the dataset and stillness of the environmental conditions, the system is able to achieve high levels of accuracy in regards to predicting the possible crops, fertilizers alongside the expected yields. Real world experiments show that it has the ability to enable efficient farming planning and decision making. SMOTE, BORUTA, and RFE are integrated for model training and efficient feature selection. Suitable for expansion, it can be adopted in combination with other frameworks to provide farmers with specific features that can boost their productivity and promote sustainability. Future upgrades will have consumers, suppliers, and buyers provide real time data, with inclusion of new crops.

REFERENCES:

[1] R. Jahan, ''Applying naive Bayes classification technique for classification of improved agricultural land soils,'' Int. J. Res. Appl. Sci. Eng. Technol., vol. 6, no. 5, pp. 189–193, May 2018.

[2] B. B. Sawicka and B. Krochmal-Marczak, ''Biotic components influencing the yield and quality of potato tubers,'' Herbalism, vol. 1, no. 3, pp. 125–136, 2017.

[3] B. Sawicka, A. H. Noaema, and A. Gáowacka, ‘‘The predicting the size of the potato acreage as a raw material for bioethanol production,’’ in Alternative Energy Sources, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: WydawnictwoNaukowe TYGIEL, 2016, pp. 158–172.

[4] B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak, ‘‘Biotic and abiotic factors influencing on the environment and growth of plants,’’ (in Polish), in Proc. BioróżnorodnośćŚrodowiskaZnaczenie, Problemy, Wyzwania. MateriałyKonferencyjne, Puławy, May 2017. [Online]. Available: https://bookcrossing.pl/ksiazka/321192

[5] R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowalski, ‘‘Response surface methodology: A retrospective and literature survey,’’ J. Qual. Technol., vol. 36, no. 1, pp. 53–77, Jan. 2004.

[6] D. K. Muriithi, ‘‘Application of response surface methodology for optimization of potato tuber yield,’’ Amer. J. Theor. Appl. Statist., vol. 4, no. 4, pp. 300–304, 2015, doi: 10.11648/j.ajtas.20150404.20.

[7] M. Marenych, O. Verevska, A. Kalinichenko, and M. Dacko, ‘‘Assessment of the impact of weather conditions on the yield of winter wheat in Ukraine in terms of regional,’’ Assoc. Agricult. Agribusiness Econ. Ann. Sci., vol. 16, no. 2, pp. 183–188, 2014.

[8] J. R. Olędzki, ‘‘The report on the state of remotesensing in Poland in 2011–2014,’’ (in Polish), Remote Sens. Environ., vol. 53, no. 2, pp. 113–174, 2015.

[9] K. Grabowska, A. Dymerska, K. Poáarska, and J. Grabowski, ‘‘Predicting of blue lupine yields based on the selected climate change scenarios,’’ ActaAgroph., vol. 23, no. 3, pp. 363–380, 2016.

[10] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, and Y. Huang, ‘‘Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning,’’ Remote Sens., vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.

[11] N. Chanamarn, K. Tamee, and P. Sittidech, ‘‘Stacking technique for academic achievement prediction,’’ in Proc. Int. Workshop Smart Info-Media Syst., 2016, pp. 14–17.

[12] W. Paja, K. Pancerz, and P. Grochowalski, ‘‘Generational feature elimination and some other ranking feature selection methods,’’ in Advances in Feature Selection for Data and Pattern Recognition, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97–112.

[13] D. C. Duro, S. E. Franklin, and M. G. Dubé, ‘‘A comparison of pixelbased and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery,’’ Remote Sens. Environ., vol. 118, pp. 259–272, Mar. 2012.

[14] S. K. Honawad, S. S. Chinchali, K. Pawar, and P. Deshpande, ‘‘Soil classification and suitable crop prediction,’’ in Proc. Nat. Conf. Comput. Biol., Commun., Data Anal. 2017, pp. 25–29.

[15] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, ‘‘Deep Gaussian process for crop yield prediction based on remote sensing data,’’ in Proc. AAAI Conf. Artif. Intell., 2017, vol. 31, no. 1, pp. 4559–4565.

[16] D. A. Reddy, B. Dadore, and A. Watekar, ‘‘Crop recommendation system to maximize crop yield in ramtek region using machine learning,’’ Int. J. Sci. Res. Sci. Technol., vol. 6, no. 1, pp. 485–489, Feb. 2019.

[17] N. Rale, R. Solanki, D. Bein, J. Andro-Vasko, and W. Bein, ‘‘Prediction of Crop Cultivation,’’ in Proc. 19th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2019, pp. 227–232.

[18] J. Jones, G. Hoogenboom, C. Porter, K. Boote, W. Batchelor, L. Hunt, P. Wilkens, U. Singh, A. Gijsman, and J. Ritchie, ‘‘The DSSAT cropping system model,’’ Eur. J. Agronomy, vol. 18, nos. 3–4, pp. 235–265, 2003.

[19] M. T. N. Fernando, L. Zubair, T. S. G. Peiris, C. S. Ranasinghe, and J. Ratnasiri, ‘‘Economic value of climate variability impact on coconut production in Sri Lanka,’’ in Proc. AIACC Working Papers, vol. 45, 2007, pp. 1–7.

[20] B. Ji, Y. Sun, S. Yang, and J. Wan, ‘‘Artificial neural networks for rice yield prediction in mountainous regions,’’ J. Agricult. Sci., vol. 145, no. 3, pp. 249–261, Jun. 2007.