

INTELLICORP AN ENSEMBLE MODEL TO PREDICT CROP USING MACHINE LEARNING ALGORITHMS

submitted in partial fulfillment of the requirements for

the award of

Bachelor of Engineering

in

Computer Science and Engineering

By

SUNNY CHOWDHARY MUPPALA

(38110304)

JACOB JASPER ROY PAGA

(38110207)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

**SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119**

MARCH 2022



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the Bonafide work of SUNNY CHOWDARY MUPPALA (38110340), JACOB JASPER ROY (38110207) who carried out the project entitled "**INTELICROP AN ENSEMBLE MODEL TO PREDICT CROP USING MACHINE LEARNING ALGORITHMS**" under my supervision from November 2021 to March 2022.

**Internal Guide
Dr. Prof. G. MEERAGANDHI**

Head of the Department

Submitted for Viva-voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I **SUNNY CHOWDHARY MUPPALA (38110340), JACOB JASPER ROY PAGA (38110207)** hereby declare that the Project Report entitled “**INTELLICORP AN ENSEMBLE MODEL TO PREDICT CROP USING MACHINE LEARNING ALGORITHMS**” done by us under the guidance of Dr.Prof.G.MEERA GANDHI (Internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering in Computer Science and Engineering.

DATE:

PLACE:

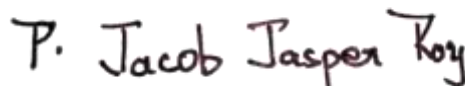
SIGNATURE OF THE CANDIDATES

1.



Sunny Chowdhary Muppala- 30110340

2.



Jacob Jasper Roy Paga- 30110207

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the Board **of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D., Dean**, School of Computing **Dr. L. Lakshmanan M.E., Ph.D.** , and **Dr. S. Vigneshwari M.E., Ph.D.** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. G. MEERAGANDHI M.E., Ph.D.** for her valuable guidance, suggestions, and constant encouragement that paved the way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

The goal of this research is to use machine learning techniques to forecast yield in different districts of India on a yearly basis. The agricultural sector is the prime occupation of India. Researchers are developing various scientific technologies in the agriculture field for better yield. In this project, we try to form an ensemble model using various machine learning algorithms for better rice production. Crop production prediction utilizing AI Strategies aims to deliver improved outcomes, but the ensemble model provides better predictive results compared to the individual algorithm. We tried to use a combination of symmetric machine learning algorithms to form an ensemble model for better prediction. Here symmetric algorithms such as random forest, Gradient Boosting, and Logistic Regression are individually used for the prediction of the yield of rice. While combining all the aforesaid algorithms to form an ensemble model of a better result. Crop recommendation depends upon soil characteristics, soil nutrient contents, topography, and other uncontrollable parameters such as weather conditions, rainfall, humidity, environment temperature. Modern technologies need to be integrated into farming practices to enhance crop productivity. The Machine Learning technique is an essential approach for building an appropriate model for crop recommendation.

TABLE OF CONTENTS

Chapter no.	Title	Pg no.
1	INTRODUCTION	1
	1.1 OUTLINE OF THE PROJECT	1
	1.2 MACHINE LEARNING	1
2	LITERATURE SURVEY	3
3	ML ALGORITHM	6
	3.1 ALGORITHM USED	6
	3.1.1 Random Forest	6
	3.1.2 Gradient Boosting	6
	3.1.3 Logistic Regression	7
4	DATASET	8
	4.1 ATTRIBUTE INFORMATION	8
	4.1.1 Summarise data	8
5	SOFTWARE REQUIREMENTS	14
	5.1 SOFTWARE USED	14
	5.1.1 Anaconda Navigator	14
	5.1.2 Python 3.7 (64-bit)	14
	5.2 LIBRARIES USED	15
	5.2.1. NumPy	15
	5.2.2. Sklenar	15
	5.2.3. Pandas	16
	5.2.4. Matplotlib	16
	5.2.5. Seaborn	17
6	METHODOLOGY	18
	6.1 WORKFLOW DIAGRAM	18

	6.1.1. Random Forest	18
	6.1.2 Logistic Regression	20
	6.1.3. Gradient Boosting	21
	6.1.4. Ensemble Model	23
7	APPLICATION WORKING	25
	7.1 Machine Learning	26
	7.2. Web Application	26
8	RESULTS AND DISCUSSION	32
9	CONCLUSION	35
	REFERENCES	36
	APPENDIX	37
	A.SOURCE CODE	37
	B.SCREENSHOTS	39
	C.PLAGIARISM REPORT	42
	D. CERTIFICATES	43
	E. JOURNAL PAPER	45
	F. COPYRIGHT FORM	51

LIST OF FIGURES

Figure no.	Title	Pg no.
4.1	Count of data points in the data variable Rainfall, N, P, K, Temperature, Humidity, and pH	9
4.2	Count of data points in the data variable N, P, K, Temperature, Humidity, and pH compared with each crop	10
4.3	Checking how (temperature, rainfall), (temperature, humidity), and (temperature, pH) affect the individual crop	12
4.4	Correlation matrix for the dataset	13
6.1	System flow diagram of the proposed method	18
6.2	Defining the random forest classifier	19
6.3	Tuning and fitting the test & train data for best estimators	19
6.4	predictions of random forest model for test and train data	20
6.5	Defining the Logistic Regression classifier	21
6.6	predictions of Logistic Regression for test and train data	21
6.7	Defining the Gradient Boosting classifier.	22
6.8	predictions of Gradient Boosting for test and train data	22
6.9	Defining the Ensemble Model classifier & voting classifier	23

6.10	prediction of ensemble model test and train data	24
7.1	Classification report of proposed ensemble model	25
7.2	Application Screenshots	29
7.3	Application Output Screenshots	30
8.1	Classification report of proposed ensemble model	34

LIST OF TABLES

Table no.	Title	Pg no.
6.1	Matrices report of proposed Random Forest	20
6.2	Matrices report of proposed Logistic Regression	21
6.3	Matrices report of proposed Gradient Boosting	23
6.4	Matrices report of proposed ensemble model	24
8.1	Accuracy of the individual machine learning algorithm and ensemble model	32

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Agriculture is the prime occupation in India. Farmers from rural areas of India are mostly dependent on agriculture for their survival. Traditional farming practices pose challenges in terms of predicting abnormal seasonal changes and foresight before flood warning signals, which results in bad crop yields. The consequence of it affects the livelihood of farmers but also badly influences the country's economic status.

Crop recommendation depends upon soil characteristics, soil nutrient contents, topography, and other uncontrollable parameters such as weather conditions, rainfall, humidity, environment temperature. Modern technologies need to be integrated into farming practices to enhance crop productivity. The Machine Learning technique is an essential approach for building an appropriate model for crop recommendation.

Farmers can't accurately predict future possibilities regarding sudden change in climatic conditions, abnormal rise in atmospheric temperature, future flood occurrences. Artificial Intelligence and Machine Learning techniques help farmers in recommending suitable crops, fertilizer recommendations, predicting the average production of the crop based on crop type. The goal of this research is to use machine learning techniques to forecast yield in different districts of India on a yearly basis.

1.2 MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The Goal of machine learning typically is to know the structure of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process

approaches. In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular variable. Thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

CHAPTER 2

LITERATURE REVIEW

- [1] (Doshi, Nadkarni, Agarwal & Shah) proposed a recommender system to make a decision about the type of crop to be grown depending upon various factors. They found the predicted result for the total amount of rainfall and combined together with the training model to recommend the type of crop to be harvested.
- The study of [2] (Mythili & Rangaraj, 2021) reveals a new deep learning-based approach for predicting crop yields. The proposed algorithm PSO-MDNN recommended a suitable crop recommender model, to recommend a suitable crop effectively. Its accuracy outperformed different machine learning techniques. The proposed model predicted yields in untested environments and recommended crop cultivation. The weight matrices with L2 regularization and PSO are used to modify the MDNN hyperparameters. The prediction accuracy was enhanced by using a network structure with optimum weights using PSO.
- [3] (Rajak, Pawar & pendke, 2017) Discuss the ensemble model in conjunction with the majority voting approach. SVM using ANN as a learner is combined with Random Tree and NB-classifier to propose a crop with higher efficiency and accuracy.
- [4] (Kuanr, Rath & Mohanty, 2018) depending on the location and agricultural preferences, proposes seed, herbicides, and equipment The agricultural yield in the Kharif season is predicted using fuzzy logic and cosine similarity. Farmers can predict crop production based on previous yield data and expected weather conditions.
- The paper [5] (Kumar, Singh, Kumar & Singh, 2015) Crop selection method (CSM) proposed and advises crop selection to enhance the net production rate.

They recommend the sort of crop to be yielded based on parameters such as rain, temperature, and humidity, as well as the type of soil.

- Another literature [6] (Pudumalar, Ramanujam, Rajashree, & et al. 2017) Precision agriculture, which makes use of soil properties, soil types, and crop production, is addressed. Based on the above characteristics, it recommends the best crop to farmers. The authors suggested an ensemble model that achieved good accuracy by combining Naive Bayes, CHAID, KNN, and Random tree.
- [7] (Lakshmi, Priya, Sahana & Manjunath, 2018) the goal was to create a crop growth monitoring recommendation system. The unique qualities utilized were soil composition, color, drainage, depth, ph. value, water holding capacity, erosion, permeability, and weather. Map-reduce with KNN suggests quickly and accurately.
- [8] (Akshatha & Shreedhara, 2018) suggested an ensemble model for precision agriculture that uses different machine learning algorithms to select the crop that is most suited to the farmer based on the information (factor influencing crop production) that he provides
- Across the US Corn Belt, homogeneous ensembles produce the most accurate yield projections. This model could accurately estimate 2019 corn grain yields with a root mean square error of 866 kg/ha, or 8.5 percent relative root mean square and could explain almost 77% of the spatiotemporal variation in corn grain yields.[9](Shahhosseini M et.al 2021)
- The authors [10](Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N.) proposed a system based on a critical evaluation of existing related studies, a prospective architecture for machine learning-based palm oil yield prediction. This system will live up to its promise by tackling new research issues in agricultural yield prediction analysis.

●[11](Thomasvan Klompenburga, and AyalewKassahuna) The most used features are temperature, rainfall, and soil type, and the most applied algorithm is Artificial Neural Networks in these models. After this observation based on the analysis of machine learning-based 50 papers, we performed an additional search in electronic databases to identify deep learning-based studies.

●[12] (Joshua, V.; Priya Harson, S.M.; Kannadasan, R.) did a statistical analysis of possible MLR approaches and machine learning algorithms such as SVM, GRNN, RBFNN, and BPNN are examined for evaluation in order to achieve improved accuracy in agricultural yield prediction. The accuracy level of the various algorithms is scrutinized using model performance indicate

CHAPTER 3

ML ALGORITHM

In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed. Two of the most widely adopted machine learning methods are: -

Supervised learning which trains algorithms based on example input and output data that is labeled by humans.

Unsupervised learning provides the algorithm with no labeled data in order to allow it to find structure within its input data.

3.1 ALGORITHM USED

Because more than one class may be given to a single instance, multi-label classification (MLC) is the best and optimum option. The MLC is embedded into the algorithms used for categorization and model building, which are as follows:

1. Random Forest
2. Gradient Boosting
3. Logistic Regression

3.1.1 Random Forest

This is an AI strategy used to predict and describe the connection between independent and dependent data values. It predicts the dependent variable by analyzing the relation between the other independent variables.

3.1.2 Gradient Boosting

This is an AI strategy that is used to solve problems like regression, classification, and other tasks. It is greedy and over fits the training dataset

quickly. The algorithm is penalized by penalizing various parts of the algorithm and the overall performance is improved.

3.1.2 Logistic Regression

This is an AI strategy used to solve problems like regression and classification. The algorithm gives an outcome that is based on the predictions of the decision trees by taking the average or mean of the output from different trees. It can be used to predict what will happen in the future.

CHAPTER 4

DATASET

The dataset consists of 22 unique crops and different values of nutrients of the soil and environmental conditions under which the crop is sustained and shown a good yield, each unique crop has 100 data conditions where they have shown better yield in different conditions of weather and soil consistency.

4.1 ATTRIBUTE INFORMATION

This dataset was created by combining various datasets of climate, rainfall, and fertilizer data for India. ICFA, India. The data consist of N, P, K, Temperature, Humidity, ph., and rainfall.

4.1.1 Summarize data

This dataset consists of various crops like Rice, Maize, Chickpea, Kidney beans, Black gram, Pigeon peas, Moth beans, Mung beans, Pomegranate, Lentil, Mango, Banana, Watermelon, Grapes, Muskmelon, Orange, Apple, Papaya, Cotton, Jute, Coconut, and Coffee. The parameters for the soil composition, weather conditions, and the count per data point are as shown in Fig. 4.1

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   N                2200 non-null   int64
1   P                2200 non-null   int64
2   K                2200 non-null   int64
3   temperature      2200 non-null   float64
4   humidity         2200 non-null   float64
5   ph               2200 non-null   float64
6   rainfall         2200 non-null   float64
7   label           2200 non-null   object
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

```
all_columns = data.columns[:-1]
```

```
plt.figure(figsize=(15,13))
```

```
i = 1
```

```
for column in all_columns[:-1]:
```

```

plt.subplot(3,3,i)
sns.histplot(data[column])
i+=1
plt.show()
sns.histplot(data[all_columns[-1]])
plt.show()

```

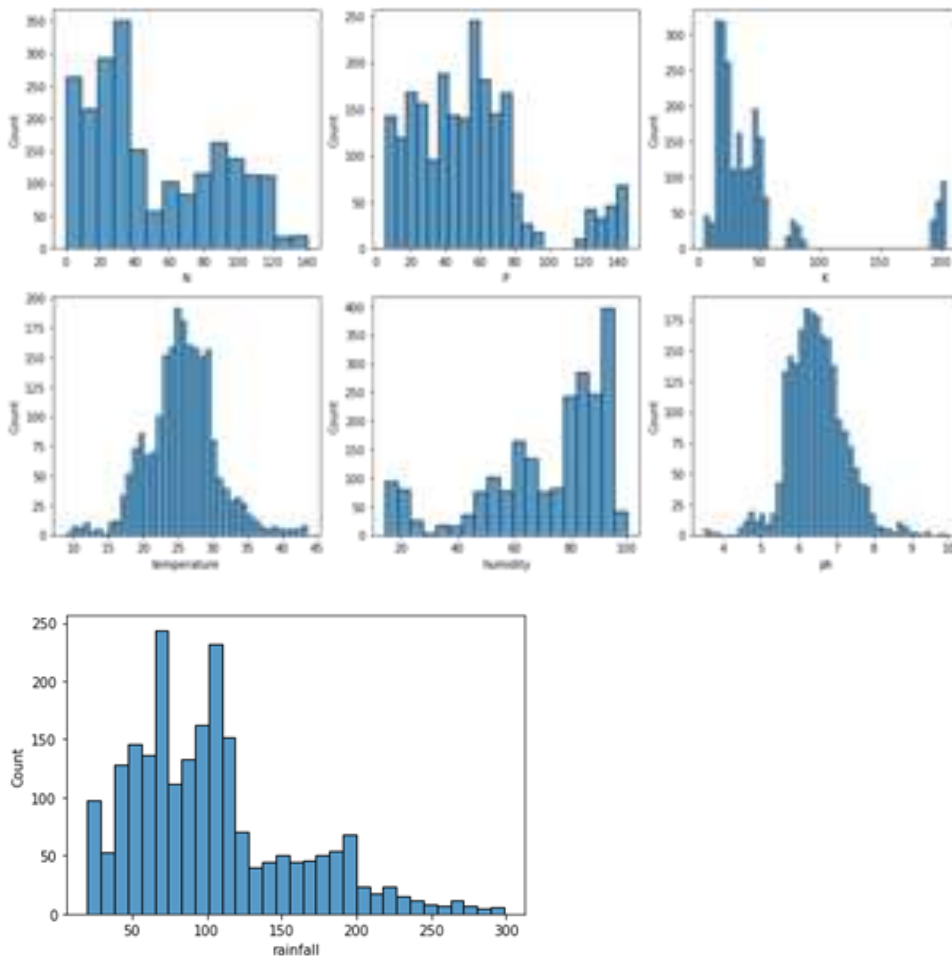


Fig. 4.1 Count of data points in the data variable Rainfall, N, P, K, Temperature, Humidity, and pH

The relation between the crop and the individual data labels are as shown in Fig. 4.2

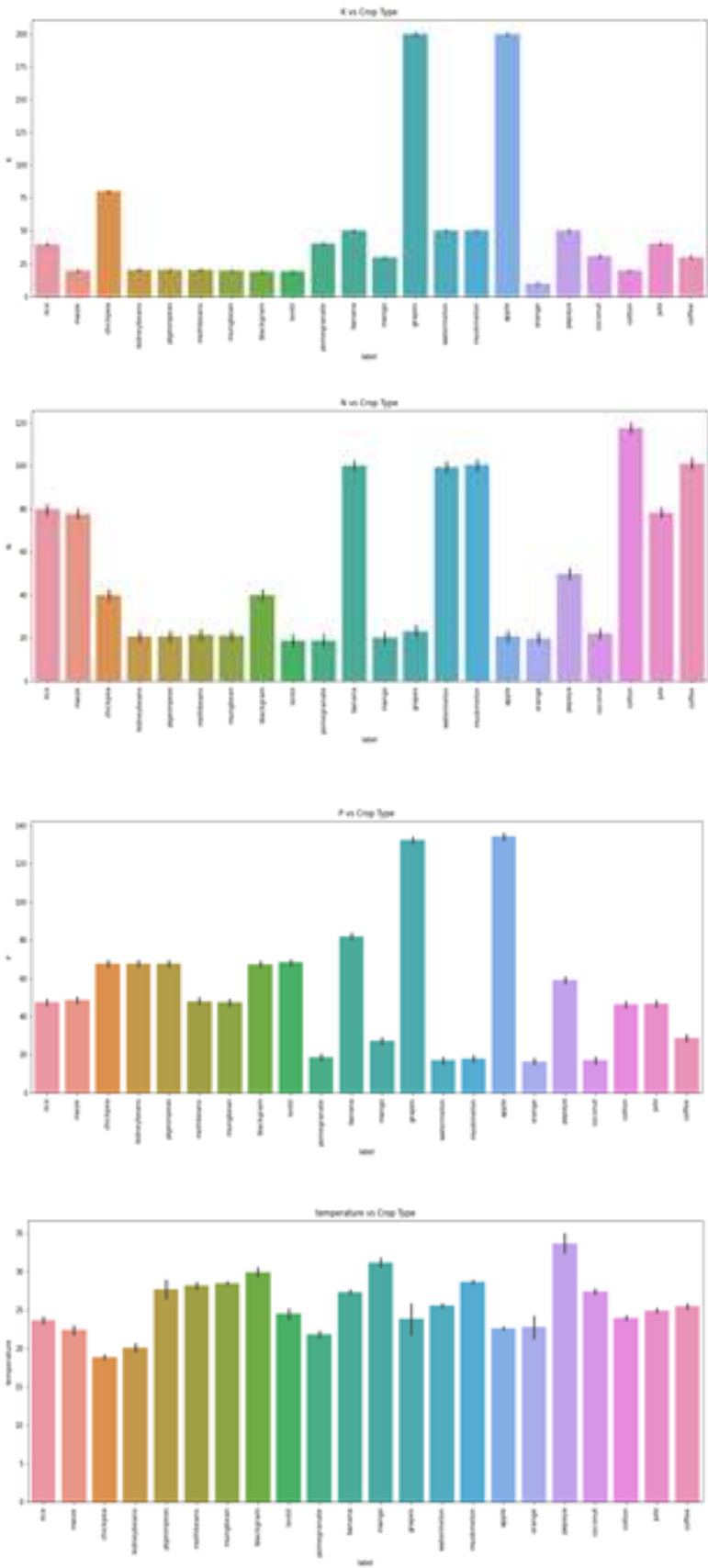
for column in all_columns:

```

plt.figure(figsize=(19,7))
sns.barplot(x = "label", y = column, data = data)
plt.xticks(rotation=90)
plt.title(f"{column} vs Crop Type")

```

plt.show()



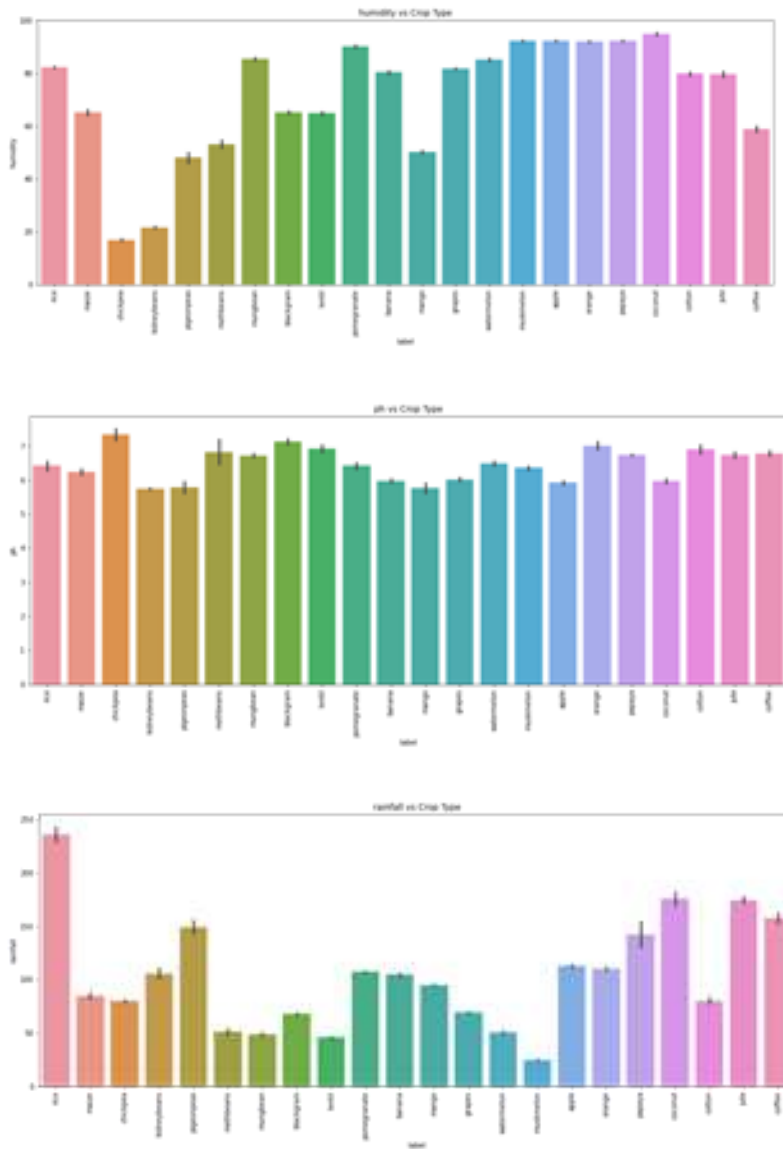


Fig. 4.2. Count of data points in the data variable N, P, K, Temperature, Humidity, and pH compared with each crop

Keeping the temperature and crops as the base and checking how the other parameters affect the crop prediction which can be seen in Fig. 4.3

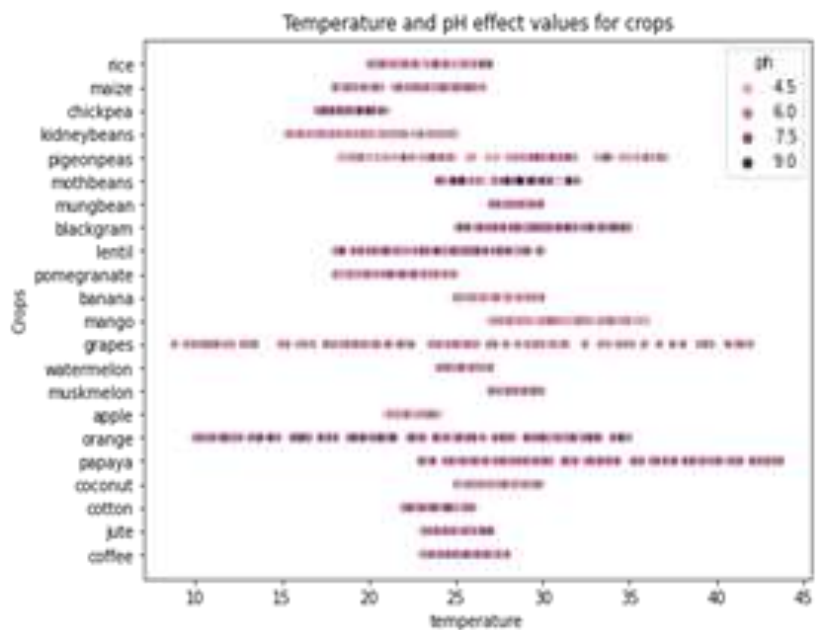
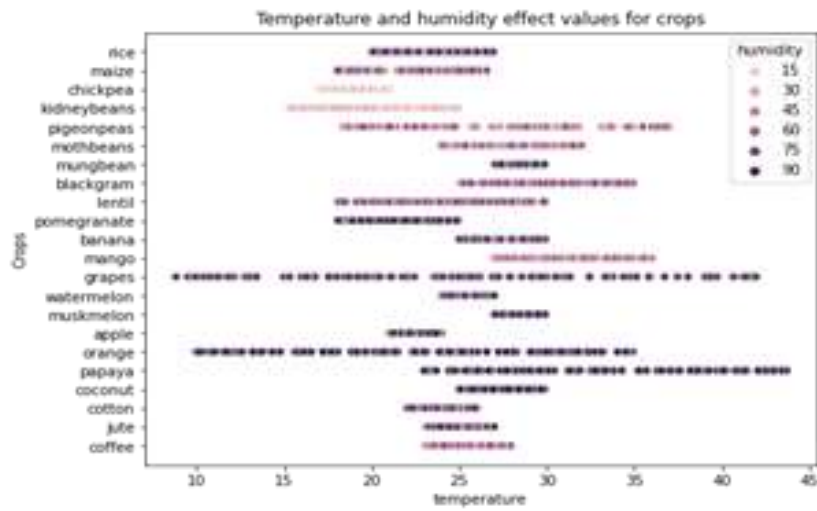
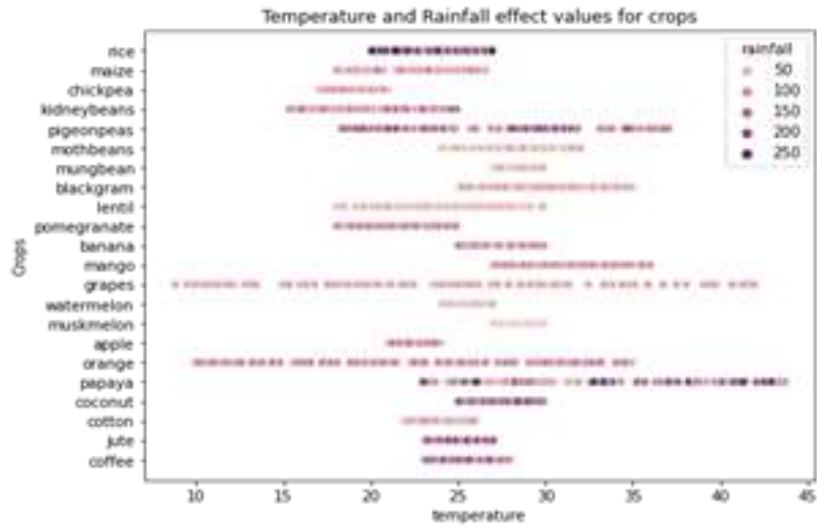


Fig. 4.3 Checking how (temperature, rainfall), (temperature, humidity), and (temperature, pH) affect the individual crop

A correlation matrix is nothing more than a table that shows the correlation coefficients for various variables. The correlation between all possible pairings of values in a table is represented by the matrix. It's a useful tool for quickly summarizing a huge dataset and identifying and visualizing trends in the data. The variables are represented by rows and columns in a correlation matrix. The correlation coefficient is contained in each cell of a table. The correlation matrix is also commonly used in combination with other forms of statistical analysis. It might be useful in the analysis of numerous linear regression models, for example. Keep in mind that the models have a number of independent variables. The correlation matrix in multiple linear regression determines the correlation coefficients between the independent variables in a model. The Correlation between all the features is as shown in Fig. 4.4

```
plt.figure(figsize=(8,8))
plt.title("Correlation between features")
corr = data.corr()
sns.heatmap(corr, annot=True)
```

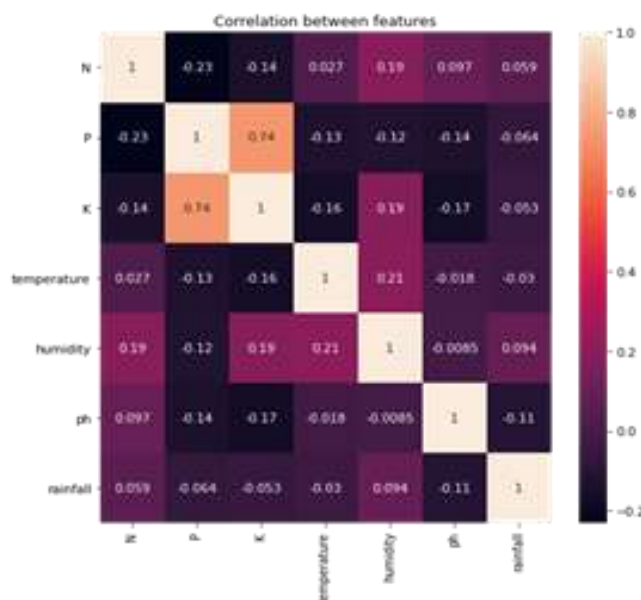


Fig 4.4. Correlation matrix for the dataset

CHAPTER 5

SOFTWARE REQUIREMENTS

5.1 SOFTWARE USED

5.1.1 Anaconda Navigator

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 15 million users and includes more than 1500 popular data-science packages suitable for Windows, Linux, and macOS We can download this from the official website of Anaconda Navigator. We will need a Jupyter Notebook, which we can directly use through Anaconda Navigator.

5.1.2 Python 3.7 (64-bit)

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

1. Open a browser window and navigate to the Download page for Windows at python.org.

2. Underneath the heading at the top that says Python Releases for Windows, click on the link for the Latest Python.
3. Release - Python 3.x.x. (As of this writing, the latest in Python 3.7.x) 3. Scroll to the bottom and select Windows x86-64 executable installer for 64-bit. We strongly recommend 64-bit because TensorFlow doesn't support 32-bit.
4. Once you have chosen and downloaded an installer, simply run it by double-clicking on the downloaded file.
5. Then just click Install Now. That should be all there is to it. A few minutes later you should have a working Python 3 installation on your system.

5.2 LIBRARIES USED

5.2.1. NumPy

To install this package with Conda run the following: `Conda install -c Conda-forge NumPy` To install this package with python, run the following: `pip install NumPy`. NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

5.2.2. Sklearn

To install this package with Conda run one of the following: `Conda install -c anaconda scikit-learn` to install this package with python, run the following: `pip install scikit-learn` Scikit-learn (formerly Scikits.learn and also known as Sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and

DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

5.2.3. Pandas

To install this package with Conda run one of the following: Conda install -c anaconda pandas to install this package with python, run the following: pip install pandas (Pandas) offer data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

5.2.4. Matplotlib

To install this package with Conda run one of the following:

Conda install -c Conda-forge matplotlib

Conda install -c Conda-forge/label/testing matplotlib

Conda install -c Conda-forge/label/testing/gcc7

matplotlib Conda install -c Conda-forge/label/gcc7

matplotlib

Conda install -c Conda-forge/label/broken

matplotlib Conda install -c Conda-forge/label/rc

matplotlib

Conda install -c Conda-forge/label/cf201901 matplotlib

To install this package with python, run the following: pip install matplotlib Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for

embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

5.2.5. Seaborn

Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions. To install this package with conda run one of the following: Conda install seaborn to install this package with python, run the following: pip install seaborn.

CHAPTER 6

METHODOLOGY

6.1 WORKFLOW DIAGRAM

The proposed method has been developed using various AI strategies like Random Forest, Logistic Regression, and Gradient Boosting combined to make an ensemble model as shown in fig.1 to predict and recommend the right crop.

Because more than one class may be given to a single instance, multi-label classification (MLC) is the best and optimum option. The MLC is embedded into the algorithms used for categorization and model building, which are as follows:

1. Random Forest
2. Logistic Regression
3. Gradient Boosting

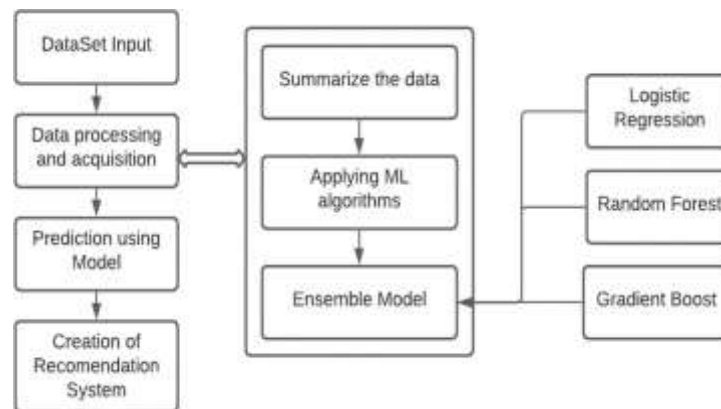


Fig. 6.1 System flow diagram of the proposed method

6.1.1. Random Forest

This is an AI strategy used to solve problems like regression and classification. The algorithm gives an outcome that is based on the predictions of

the decision trees by taking the average or mean of the output from different trees. It can be used to predict what will happen in the future.

We have defined the random forest classifier and then we have tuned the parameters with the help of grid search and then with the tuned parameters, we have fit the train and test data which can be seen in Fig. 6.2

```
rf = RandomForestClassifier()

params_rf = {'n_estimators': [50, 100, 200]}

rf_gs = GridSearchCV(rf, params_rf, cv=5)

rf_gs.fit(X_train, y_train)

GridSearchCV(cv=5, estimator=RandomForestClassifier(),
             param_grid={'n_estimators': [50, 100, 200]})
```

Fig. 6.2 Defining the random forest classifier

After tuning the parameters and fitting the test and train data we performed the best estimator and we got the result as in Fig. 6.3 and Fig 6.4

```
[ ] rf_best = rf_gs.best_estimator_

print(rf_gs.best_params_)

{'n_estimators': 200}
```

Fig. 6.3. Tuning and fitting the test & training data for best estimators

```
[ ] y_pred = rf_best.predict(X_test)

acc = accuracy_score(y_test, y_pred)

print("rf aproach")
print("MAE:\t",metrics.mean_absolute_error(y_test,y_pred))

print("MSE:\t",metrics.mean_squared_error(y_test,y_pred))

print("RMSLE:\t",np.sqrt(metrics.mean_squared_log_error(y_test,y_pred)))

print('knn: {}'.format(rf_best.score(X_test, y_test)))

rf aproach
MAE:      0.06666666666666667
MSE:      0.7424242424242424
RMSLE:    0.09318123679872903
knn: 0.9939393939393939
```

Fig. 6.4 prediction of random forest model for test and train data

Here is the table of metrics for the particular machine learning algorithm which is in Table 6.1.

Table 6.1. Matrices report of proposed Random Forest

MAE	MSE	RMSLE	Accuracy
0.0666	0.7424	0.0931	0.9939

6.1.2 Logistic Regression

This is an AI strategy used to predict and describe the connection between independent and dependent data values. It predicts the dependent variable by analyzing the relation between the other independent variables.

We have defined the Logistic Regression classifier and then we have tuned the parameters with the help of grid search and then with the tuned parameters, we have fit the train and test data which can be seen in Fig. 6.5.

```

▶ log_reg = LogisticRegression(C=0.1, max_iter=100000)

log_reg.fit(X_train, y_train)

LogisticRegression(C=0.1, max_iter=100000)

```

Fig. 6.5 Defining the Logistic Regression classifier

After tuning the parameters and fitting the test and train data we performed the best estimator and we got the result as in Fig. 6.6.

```

[ ] y_pred = log_reg.predict(X_test)

acc = accuracy_score(y_test, y_pred)

print("log_reg approach")
print("MAE:\t",metrics.mean_absolute_error(y_test,y_pred))

print("MSE:\t",metrics.mean_squared_error(y_test,y_pred))

print("RMSLE:\t",np.sqrt(metrics.mean_squared_log_error(y_test,y_pred)))

print('knn: {}'.format(log_reg.score(X_test, y_test)))

log_reg approach
MAE:      0.15151515151515152
MSE:      1.6303030303030304
RMSLE:    0.0913287976488097
knn: 0.9833333333333333

```

fig. 6.6 prediction of Logistic Regression for test and train data

Here is the table of metrics for the particular machine learning algorithm which is in Table 6.2

Table 6.2. Matrices report of proposed Logistic Regression

MAE	MSE	RMSLE	Accuracy
0.1515	1.6303	0.0913	0.9833

6.1.3. Gradient Boosting

This is an AI strategy that is used to solve problems like regression, classification, and other tasks. It is greedy and over lifts the training dataset

quickly. The algorithm is penalized by penalizing various parts of the algorithm and the overall performance is improved.

We have defined the Gradient Boosting classifier and then we have tuned the parameters with the help of grid search and then with the tuned parameters, we have fit the train and test data which is visible in Fig. 6.7.

```
[ ] gbr = GradientBoostingClassifier(n_estimators=20, max_depth=4, max_features=2, random_state=0)
    gbr.fit(X_train, y_train)

GradientBoostingClassifier(max_depth=4, max_features=2, n_estimators=20,
                           random_state=0)
```

Fig. 6.7 Defining the Gradient Boosting classifier

After tuning the parameters and fitting the test and train data we performed the best estimator and we got the result as in Fig. 6.8.

```
[ ] y_pred = gbr.predict(X_test)

acc = accuracy_score(y_test, y_pred)

print("grb aproach")
print("MAE:\t",metrics.mean_absolute_error(y_test,y_pred))

print("MSE:\t",metrics.mean_squared_error(y_test,y_pred))

print("RMSLE:\t",np.sqrt(metrics.mean_squared_log_error(y_test,y_pred)))

print('grb: {}'.format(gbr.score(X_test, y_test)))

grb aproach
MAE:      0.11515151515151516
MSE:      1.1575757575757575
RMSLE:    0.07636804326905322
```

Fig. 6.8 prediction of Gradient Boosting for test and train data

Here is the table of metrics for the particular machine learning algorithm which is in **Table 6.3**.

Table 6.3. Matrices report of proposed Gradient Boosting

MAE	MSE	RMSLE	Accuracy
0.1151	1.1575	0.0763	0.9863

6.1.4. Ensemble Model

We have predicted the accuracy with the created ensemble model and we acquired an accuracy of 99.54%. To reach this high accuracy we must tune the hyperparameters for each model to find the optimal parameters to get this result and ensemble it into a single model to make predictions a bit easier.

We have defined the Ensemble Model classifier and then we have tuned the parameters with the help of grid search and then with the tuned parameters, we have fit the train and test data which can be seen in Fig. 6.9.

```

from sklearn.ensemble import VotingClassifier

estimators=[('grb', gbr), ('rf', rf_best), ('log_reg', log_reg)]

ensemble = VotingClassifier(estimators, voting='hard')

[ ] ensemble.fit(X_train, y_train)

ensemble.score(X_test, y_test)

```

Fig:6.9. Defining the Ensemble Model classifier & voting classifier

After tuning the parameters and fitting the test and train data we performed the best estimator and we got the result as in Fig. 6.10.

```
[ ] y_pred = ensemble.predict(X_test)

acc = accuracy_score(y_test, y_pred)

print("ensemble aproach")
print("MAE:\t",metrics.mean_absolute_error(y_test,y_pred))

print("MSE:\t",metrics.mean_squared_error(y_test,y_pred))

print("RMSLE:\t",np.sqrt(metrics.mean_squared_log_error(y_test,y_pred)))

print('ensemble: {}'.format(ensemble.score(X_test, y_test)))

ensemble aproach
MAE:      0.05303030303030303
MSE:      0.6196969696969697
RMSLE:    0.07596648236789506
ensemble: 0.9954545454545455
```

Fig. 6.10 prediction of ensemble model test and train data

The matrix for the created ensemble model is as follows which is in **Table 6.4.**

Table 6.4. Matrices report of proposed ensemble model

MAE	MSE	RMSLE	Accuracy
0.0530	0.6196	0.0759	0.9954

CHAPTER 9

CONCLUSION

In this study, we developed a crop recommendation algorithm that farmers in India may utilize. The suggestion method will assist farmers in obtaining the optimum crop to plant for the given geographical circumstances. The model developed in this study seeks to provide the optimum crop suggestion based on the user's geographical conditions with an accuracy of 99.54 percent. The proposed approach might be highly beneficial to farmers since it provides precise findings, which means there is a strong probability of maximizing production by depending on the conditions specified by the soil test and the meteorological conditions in the region.

REFERENCES

- [1] Doshi, Zeel et al. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (2018): 1-6.
- [2] Mythili K., Rangaraj R. 2021. "Crop Recommendation for Better Crop Yield for Precision Agriculture Using Ant Colony Optimization With Deep Learning Method". Annals of the Romanian Society for Cell Biology, April, 4783-94. <https://www.annalsofscb.ro/index.php/journal/article/view/3024>.
- [3] Rajak, R. K., Pawar, A., Pendke, M., & et al. (2017). "Crop recommendation system to maximize crop yield using machine learning technique." International Research Journal of Engineering and Technology (IRJET), 4(12), 950-953.
- [4] Kuanr, Madhusree & Rath, Bikram & Mohanty, Sachi. (2018). "Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model." International Journal of Engineering & Technology. 7. 277. 10.14419/ijet.v7i4.15.23006.
- [5] Kumar, R., Singh, M. P., Kumar, P. & Singh, J.P. (2015). "Crop selection method to maximize crop yield rate using machine learning technique." 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy, and Materials (ICSTM), 138-145.
- [6] Pudumalar, S. et al. "Crop recommendation system for precision agriculture." 2016 Eighth International Conference on Advanced Computing (ICoAC) (2017): 32-36.
- [7] Bandara, Pradeepa & Weerasooriya, Thilini & T.H., Ruchirawya & Nanayakkara, W.J.M. & M.A.C, Dimantha & M.G.P, Pabasara. (2020). "Crop Recommendation System." International Journal of Computer Applications. 175. 22-25. 10.5120/ijca2020920723.
- [8] Akshatha. K. R., & Shreedhara. K.S. (2018). "Implementation of machine learning algorithms for crop recommendation using precision agriculture." International Journal of Research in Engineering, Science and Management (IJRESM), 1(6), 58-60.

A. PLAGIARISM REPORT

InteliCrop Paper Final

ORIGINALITY REPORT

3%

SIMILARITY INDEX

1%

INTERNET SOURCES

2%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

Balaji Prabhu B.V., M. Dakshayini. "An Effective Multiple Linear Regression-Based Forecasting Model for Demand-Based Constructive Farming", International Journal of Web-Based Learning and Teaching Technologies, 2020

Publication

1%

2

indjst.org
Internet Source

1%

3

"Crop Recommendation using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering, 2019

Publication

<1%

4

Chouaib El Hachimi, Salwa Belaqqiz, Saidi Khabba, Abdelghani Chehbouni. "Towards precision agriculture in Morocco: A machine learning approach for recommending crops and forecasting weather", 2021 International Conference on Digital Age & Technological

<1%

B. CERTIFICATES





St. JOSEPH'S COLLEGE OF ENGINEERING

(An Autonomous Institution)

St. Joseph's Group of Institutions

Jeppiaar Educational Trust

OMR, Chennai - 600 119, Tamil Nadu, India.

CERTIFICATE

Sponsored by



ACCAI 217

Peer Reviewed

This certificate is presented to



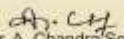
Jacob Jasper Roy Paga

Research Scholar

Sathyabama Institute of Science and Technology

Chennai, Tamil Nadu, India.

for presenting the research paper entitled "InteliCrop: An Ensemble Model to Predict Crop using Machine Learning Algorithms" in the IEEE sponsored 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI 2022) held at the St. Joseph's College of Engineering, Chennai, Tamil Nadu, India, during 28 - 29, January 2022.


Dr. A. Chandra Sekar
Conference Chair


Dr. Vaddi Sessa Giri Rao
Principal



2022 IEEE International Conference on
**Advances in Computing, Communication
and Applied Informatics (ACCAI 2022)**

28th - 29th January, 2022 | Chennai | India

COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

InteliCrop: An Ensemble Model to Predict Crop using Machine Learning Algorithms

Muppala Sunny Chowdhary, Jacob Jaspe Roy Paga, Sasmita S. Choudhury, Sachi Nandan Mohanty, Meera Gandhi

2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)

COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the [IEEE PSPB Operations Manual](#).
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

You have indicated that you DO wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."

CONSENT AND RELEASE

1. In the event the author makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the author, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.
2. In connection with the permission granted in Section 1, the author hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

Sunny Chowdhary Muppala

06-01-2022

Signature

Date (dd-mm-yyyy)

Information for Authors

AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightrresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use. The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

AUTHOR ONLINE USE

- **Personal Servers.** Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
- **Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.
- **Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the

IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

Questions about the submission of the form or manuscript must be sent to the publication's editor.

Please direct all questions about IEEE copyright policy to:

IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966

