E-Farm Crop Prediction Using Machine Learning

Sangeetha Sathyan
Department of Computer Application
Amal Jyothi College of Engineering Kanjirappally, India
sangeethasathyan2022b@mca.ajce.in

Mr. T.J Jobin
Department of Computer Application
Amal Jyothi College of Engineering Kanjirappally, India
tjjobin@amaljyothi.ac.in

***Abstract* — We all know that agriculture is important to the majority of Indians. Farmers nowadays are curious about crop yields when seasons, ambient temperature, rain, and soil conditions change. Farmers can grow new types of crops, which supports the prediction and boosts the crop's profit. Farmers will be aided in picking which crop to plant in order to maximise yield by machine learning algorithms' forecasts, which will take into consideration elements such as temperature, rainfall, and area. Crop yield prediction in Python is the best and easiest way to improve accuracy by include N, P, K, and soil type as attributes. We're trying to figure out the best and most accurate way to anticipate crop yields . The ML Algorithms like SVM and Random Forest Regressor,XGB regressor are used to predict crop yield.**

Keywords—Crop Yield ,Prediction,ML Algorithms, SVM,Random Forest Regressor,XGB Regressor

I. INTRODUCTION

In India agriculture is one amongst the foremost important occupations. . It is the fastest-growing economic sector in the country and is crucial to its overall development. The importance of yield prediction in agriculture cannot be overstated. Farmers used to be able to predict their harvest based on the yield of the previous year. Thus, there are various strategies or algorithms for this type of knowledge analytics in crop prediction, and we will predict crop production with the use of these algorithms. It uses the random forest algorithm, polynomial SVM, XGB Regressor, and neural network.

Crop seasons, crop year,pressure,temperature, Humidity, and crop name are some of the weather parameters and soil composition that the built system will consider when recommending the best crop for a specific agricultural situation. .Temperature, humidity, and pressure data are collected from farmers or sensors by the system. For the farmer, the approach offers a crop and the amount of nutrients that should be added for the projected crop.

The method for crop yield prediction that uses:

- Load Dataset..
- Data Processing: It contain Train and Test split data
- Apply Machine Learning Algorithm for crop prediction(SVM,Random Forest Regressor , Neural network ,XGB Regressor).

II. LITERATURE REVIEW

Ashwani kumar Kushwaha[1] Describes crop yield prediction methodologies and offers a suitable crop to increase farmer profits and the quality of the agriculture industry..This study uses Hadoop platform and agro algorithm to acquire huge volume data, also known as big data (soil and meteorological data), for crop yield prediction. As a result of the repository data, crop appropriateness for a specific situation may be predicted, and crop quality can be improved.

Girish L [2] uses a machine learning algorithm to estimate crop productivity and rain precipitation. They went over many machine learning methodologies for predicting rainfall and crop yield in this research, as well as the efficiency of various machine learning algorithms such as liner regression, SVM, KNN method, and decision tree. They conclude that SVM has the highest effectiveness for rainfall prediction in that method.

Rahul Katarya [3] discusses the many machine learning technologies that can be utilised to increase crop productivity. They went over many artificial intelligence strategies in this paper, for precision agriculture, such as machine learning methods and huge data analysis Crop recommender systems such as KNN, Ensemble-based Models, and others are discussed.

Mishra [4] has theoretically described a number of machine learning approaches that can be used in a variety of forecasting applications. However, because their study does not include any algorithms, it is unable to provide a clear picture of the suggested work's feasibility.

Dr. Y. Jeevan Nagendra Kumar [5] discovered that Machine Learning algorithms can predict a target/outcome using Supervised Learning. Agricultural yield prediction using supervised learning algorithms is the subject of this study. According to the study, the forecasts might be made using the Random Forest ML approach, which claims to produce the best accurate crop forecast with the fewest number of models.

III. MOTIVATION

Crop prediction could be a common occurrence. During the growing season, a farmer was curious to know what percentage yield he should expect. In the past, yield prediction was based on the farmer going straight to Instead than focusing about crop forecasting, the present approach focuses on yield prediction. Unless the right crop is forecast, the yield will be improved, and existing pesticide systems will be used, environmental and meteorological parameters linked with the crop will be ignored.

DOI: 10.5281/zenodo.6907185
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

One of the most important conditions for agricultural progress is to promote and ease agricultural production at a faster rate.One of the most important conditions for agricultural progress is to promote and ease agricultural production at a faster rate.

Any crop's production will guide the way either through domain interest or yield enhancement, or both. As a result, changes in crop productivity continue to vex the realm and cause severe distress. As a result, in order to overcome the current challenge, good crop forecast techniques must be tried.

This model will help the farmers to simply predict the crop yeild during this system users are employing a dataset to create a model. The model is trained using this dataset build using support vector machine. Then the model is employed to predict the yield.

IV. METHODOLOGY

The purpose of this paper is to search out a Machine Learning approach to spot crop yield prediction using the given dataset with accuracy. The dataset is split into two training and testing data sets. Dataset named finalized_dataset.csv contain Maharashtra State parameters in India, because the climate changes from place to place, it absolutely was necessary to induce data at district level. Historical data about the crop and also the climate of a particular region was needed to implement the system. Main Methodologies are used for feature selection polynomial SVM,Random Forest,XGB Regressor. The structure is

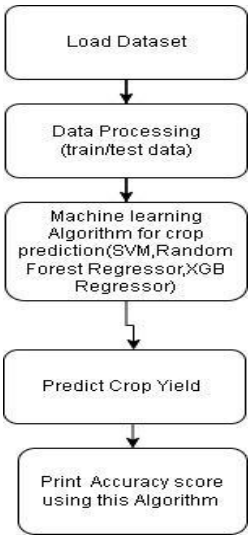


Figure 1: E-Fram Crop Yield Prediction method. The

steps that require to be followed are:

1. Load Dataset
2. Data Pre-processing
3. Machine learning algorithm for crop prediction
4. Predict crop yield
5. Print Accuracy score

A. **Load Dataset:** User have a dataset for crop yield prediction finalized_dataset.csv, the attributes of this dataset is used to train the model. There 7255 rows and 85 columns in dataset Parameters arecrop_Year,season , crop name,temperature ,Ph,Pressure…tec

B. **Data Pre-processing:**In ML, data preparation may be a critical step that helps improve the level of knowledge and pushes the extraction of relevant insights from the data. The process of preparing (cleaning and organising) data so that it can be used to create and train Machine Learning models.

C. **Machine Learning:**This data is used to forecast agricultural yields. SVM, Random Forest, and XGB Regressor neural network implementation are examples of machine learning algorithms that deliver high accuracy.

1. **Polynomial SVM[2]-** Is a kernel function that allows learning of non-linear models by reflecting the similarity of vectors (training samples) over a feature space over polynomials of the initial variables. Support vector machines (SVMs) and other kernelized models are frequently employed with it.
2. **Random Regressor[5] –**.A random forest could be a meta estimator that matches a variety of classification decision trees on various subsamples of the dataset and uses the average to increase predictive accuracy and control overfitting.
3. **XGB Regressor-**Its stands Extreme Gradient Booster The results of regression problems are continuous or real values. Some commonly used regression algorithms are regression toward the mean and decision trees. Several measures are involved within the regression, like root mean square error (RMSE) and root mean square error (MAE). These are some key members of the XGBoost models, all playing a vital role.
4. **Neural network Implementation-** A neural network may be a system that learns the way to make predictions by taking the dataset,Making a accurate prediction Comparing the all prediction to the specified predicted output and adjusting its internal state to predict correctly the following time.

V.BUILD MODEL

The model building is the main step in the crop yield prediction. While building the model user use the algorithms. The steps involved are:

1. Import the packages that are necessary

```
[ ] import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as seabornInstance
    from sklearn.linear_model import LinearRegression
    from sklearn import metrics
    %matplotlib inline

[ ] from sklearn import preprocessing

[ ] from sklearn.preprocessing import StandardScaler
```

2. Add the data into a DataFrame, then get the shape

DOI: 10.5281/zenodo.6907185
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

of data.

```
import pandas as pd
df = pd.read_csv('/content/drive/MyDrive/crop yield/finalised_dataset.csv',na_values='=')
df
```

	Unnamed: 0	state_names	district_names	crop_year	season_names	crop_names	area	temperature	wind_speed	pressure	humidity	soil_type	N
0	125191	Maharashtra	AHMEDNAGAR	1997	Autumn	Maize	1.0	20.770894	2.068260	1014.863796	21.947147	loamy	56.070
1	125192	Maharashtra	AHMEDNAGAR	1997	Kharif	ArhariTur	17600.0	20.160426	1.976480	1015.193528	20.643237	sandy	9.000
2	125193	Maharashtra	AHMEDNAGAR	1997	Kharif	Bajra	274100.0	21.998299	2.000534	1014.184607	21.422312	clay	0.000
3	125194	Maharashtra	AHMEDNAGAR	1997	Kharif	Gram	40800.0	21.776377	2.019750	1015.053118	21.810567	chalky	38.250
4	125195	Maharashtra	AHMEDNAGAR	1997	Kharif	Jowar	900.0	20.075734	1.974351	1015.170230	21.930206	clay	0.000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
12623	137814	Maharashtra	YAVATMAL	2014	Rabi	Jowar	4000.0	21.635879	2.000060	1014.302213	20.060662	silty	7.840
12624	137815	Maharashtra	YAVATMAL	2014	Rabi	Maize	1300.0	21.709611	2.053609	1015.803912	21.263478	silt	7.504
12625	137816	Maharashtra	YAVATMAL	2014	Rabi	Wheat	29100.0	21.851730	2.027476	1014.031903	20.059945	sandy	0.000
12626	137817	Maharashtra	YAVATMAL	2014	Summer	Groundnut	9400.0	21.569380	2.004421	1013.989126	21.835159	chalky	2.896

3. Then split the dataset into training and testing datasets.

```
a_train, a_test, b_train, b_test = train_test_split(a, b, test_size = 0.3, random_state = 42)
```

4. Initialize random forest regressor and calculate the accuracy

```
[ ]   from sklearn.ensemble import RandomForestRegressor
      regr = RandomForestRegressor(max_depth=2, random_state=0, n_estimators=10
      regr.fit(a_train, b_train)
      b_pred = regr.predict(a_test)
from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error as mae
from sklearn.metrics import r2_score

print('MSE =', mse(b_pred, b_test))
print('MAE =', mae(b_pred, b_test))
print('Random Forest Regressor Accuracy Score =', r2_score(b_pred, b_test))
```

The accuracy is 95%

5. Initialize polynomial SVM

```
[ ] from xgboost import XGBRegressor
    from sklearn.metrics import mean_absolute_error
    XGBModel = XGBRegressor()
    XGBModel.fit(a_train,b_train , verbose=False)
```

7. Then identify on test set and calculate the accuracy of the model.

```
XGBpredictions = XGBModel.predict(a_test)
MAE = mean_absolute_error(b_test , XGBpredictions)
print('XGBoost validation MAE = ',MAE)
XGBpredictions
```

```
    print('MSE =', mse(b_pred, b_test))
    print('MAE =', mae(b_pred, b_test))
    print('R2 Score =', r2_score(b_pred, b_test))
```

The accuracy is 98.4%.

8. Print the correlation heat map

9. Implement the neural network and plot the mean absolute error and epoch

```
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Activation, Flatten
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from matplotlib import pyplot as plt
import seaborn as sb
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import warnings
from tensorflow.keras.callbacks import History
warnings.filterwarnings('ignore')
warnings.filterwarnings('ignore', category=DeprecationWarning)
```

```
NN_model = Sequential()

# The Input Layer :
NN_model.add(Dense(128, kernel_initializer='normal',input_dim = a_train.shape[1], activation='relu'))

# The Hidden Layers :
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))
NN_model.add(Dense(256, kernel_initializer='normal',activation='relu'))

# The Output Layer :
NN_model.add(Dense(1, kernel_initializer='normal',activation='linear'))
```

V1. RESULT

The result shows that the crop yield prediction based on the dataset given. The Polynomial SVM,XGBregressor,neural network and Random Forest regressor provides correct result with high accuracy. The dataset crop.csv contain crop from Maharastra state contains seasons,weather conditions,temperature,pressure etc.that provide correct yield.

*Yeild is predicted

```
df['Yield'] = df['production']/df['area']
df
```

ames	crop_names	area	temperature	wind_speed	pressure	...	other oilseeds	Maharashtra	chalky	clay	loamy	peaty	sandy	silt	silty	Yield
harif	ArhariTur	12200.0	20.768143	2.002031	1013.280471	...	0	1	0	0	0	1	0	0	0	0.393443
harif	Bajra	240500.0	20.722713	2.105239	1015.061641	...	0	1	0	0	0	0	0	0	1	0.596258
harif	Groundnut	5300.0	21.419190	2.046843	1015.770055	...	0	1	0	0	0	1	0	0	0	0.735849
harif	Jowar	100.0	20.425919	2.024060	1013.971163	...	0	1	0	0	0	0	1	0	0	1.000000
harif	Maize	11400.0	20.823344	1.989898	1015.453191	...	0	1	0	0	0	0	0	0	1	1.228070
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Rabi	Jowar	4000.0	21.635879	2.000060	1014.302213	...	0	1	0	0	0	0	0	0	1	0.825000
Rabi	Maize	1300.0	21.709611	2.053609	1015.803912	...	0	1	0	0	0	0	0	1	0	0.153846
Rabi	Wheat	29100.0	21.851730	2.027476	1014.031903	...	0	1	0	0	0	0	1	0	0	0.920962
mmer	Groundnut	9400.0	21.569380	2.004421	1013.989126	...	0	1	1	0	0	0	0	0	0	1.223404

DOI: 10.5281/zenodo.6907185
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

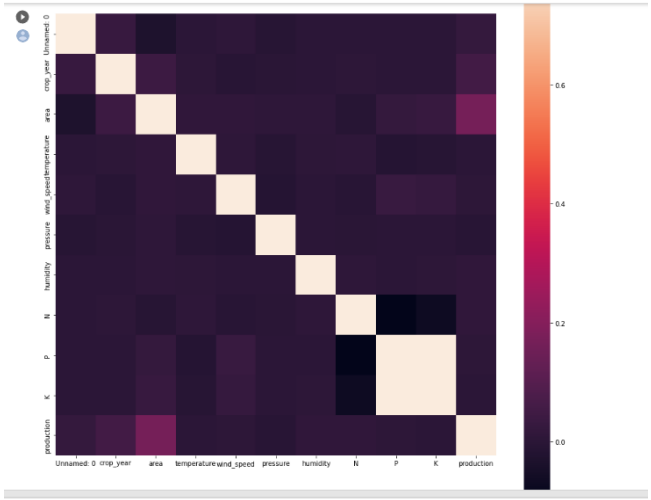
* Corelation HeatMap for Yield.
* Accuracy Score for Random Regressor and XGB regressor

*Neural network implementation with train and test data

```
MSE = 7.671048879964048
MAE = 0.8953650873829122
R2 Score = 0.9589614680509005
```

```
print(r2_score(b_test , XGBpredictions))
```

```
0.9654928330252374
```



V11.CONCLUSION

This paper provides a method for crop yield prediction using machine learning algorithms. It's all about using the best features to get the best accuracy score. Here, the user implemented a solution by first processing the data, then splitting the training and testing data to derive some features of the dataset. The user has then used the SVM algorithm, XGB Booster classifier, Random Forest classifiers and neural network implementation to build a model to predict performance.

The algorithm provides the best accuracy comparison with existing algorithms. Provides the yield rate of the crop in the final prediction.

The other scope is that the key sources of crop prediction can be found using the correlation heatmap.

REFERENCES

[1] Ashwani kumar Kushwaha, Swetabhattachrya "crop yield prediction using agro algorithm in hatoop".

[2] Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T "Crop Yield and Rainfall Prediction in Tumakuru District using Machine.

[3] Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper "Impact of Machine Learning Techniques in Precision Agriculture".

[4] Prof. D.S. Zingade ,Omkar Buchade ,Nilesh Mehta ,Shubham Ghodekar ,Chandan Mehta "Crop Prediction System using Machine Learning".

[5] Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja "Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine

DOI: 10.5281/zenodo.6907185
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam