Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 4661–4668, DOI 10.12694/scpe.v25i6.3304

PREDICTIVE CULTIVATION: INTEGRATING METEOROLOGICAL DATA AND MACHINE LEARNING FOR ENHANCED CROP YIELD FORECAST

BJD KALYANI[*], SHAIK SHAHANAZ[†] AND KOPPARTHI PRANEETH SAI[‡]

Abstract. Agriculture is a key component of Telangana's economy, and greater performance in this sector is crucial for inclusive growth. A central challenge is yielding estimation to predict crop yields before harvesting. This paper addresses this challenge with machine learning approaches includes Naive Bayes, KNN and Random Forest. The parameters considered for model testing are crop, season, rainfall and location. This paper includes a case study of Telangana with the help of Telangana weather data set to provide analysis on the key factors like overall rainfall recorded with respect to each Mandal, overall seasonal yield in selected years, seasonal yield of major crops like Bengal gram, groundnut and maize, and overall yield in two different agricultural seasons: rabi and kharif. Random forest machine learning model produces highest accuracy of 99.32% when compared with other process models.

Key words: Prediction Cultivation, Machine Learning, Smart Agriculture, Random Forest, Meteorological data

1. Introduction. A harvest expectation is a boundless issue that happens. During the rising season, a rancher had an interest in knowing how much yield he is going to anticipate. In the prior period, this yield forecast becomes a self-evident truth depended on Farmer's drawn-out understanding for explicit yield, crops and climatic conditions [1]. Rancher legitimately goes for yield forecast instead of worried on crop expectation with the current framework [2]. Except if the right harvest is anticipated how the yield will be better and also with existing frameworks pesticides, natural and meteorological parameter [3] identified with the crop isn't thought of. Advancing and alleviating the rural creation at an all the more quickly pace is one of the fundamental circumstances for farming improvement. Any harvest's creation shows the route either by the enthusiasm of area or improvement in yield or both.

In India, the possibility of augmenting the locale under any yield doesn't exist with the exception of by restoring to increment trimming quality or harvest substitution. Along these lines, varieties in a difficult situation the region and create thorough trouble. In this way, there is have to endeavour great procedure for crop expectation so as to conquer the existing issue. The objective of this paper is to develop an application using Machine Learning for Predicting which Crop yield based on Meteorological data using "K nearest neighbour classification" (KNN) [4], Naive Bayes [5] and Random Forest [6].

2. Related Work. Meteorological data have been heavily utilised by the remote sensing [7] group to forecast agricultural productivity. However, all of the strategies rely on hand-crafted features, assuming that they can effectively capture the majority of the vegetation growth [8] information offered in high-dimensional images. Elavarasan et al., [9] concentrates on crop yield prediction rely on climatic parameters and focus to identify more parameters resulting high crop yield. Chlingaryan and Sukkarieh et al., [10] carried out survey on crop yield prediction with crop, water, soil and livestock management based on machine learning models. Liakos et al., [11] illustrates cost effective machine learning solutions integrating with remote sensing technology for efficient crop yield prediction. Balamurugan et al., [12] focus on integration of various parameters includes rainfall, temperature and season with random forest classifier. Aruvansh Nigam, Saksham Garg, Archit Agrawal et al., [13] demonstrates crop yield prediction with various algorithms includes random forest machine learning algorithm for crop yield prediction, recurrent neural network for rainfall prediction and LSTM for temperature

[*]Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Telangana, India (kjd_kalyani@yahoo.co.in)
[†]Department of Computer Science and Engineering, Vardaman College of Engineering, India
[‡]Department of Computer Science, Lamer University, USA.

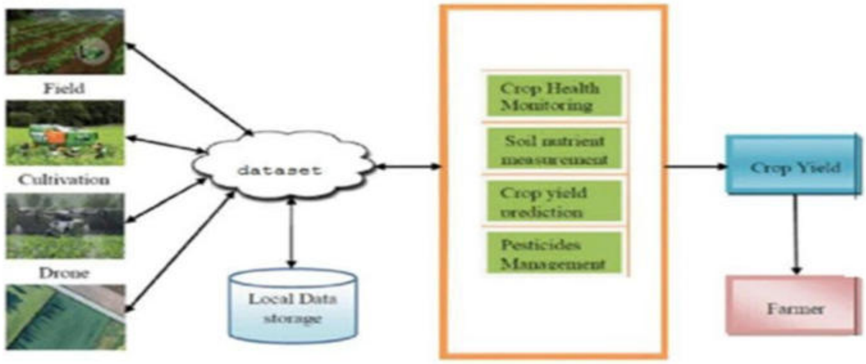

Fig. 2.1: Proposed System Architecture

Table 3.1: Sample Dataset

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 8 columns

prediction. Thus, the literature focus on limited set of crop pictures considered for prediction. The proposed methodology can integrate computer vision [14] and Machine learning strategies with the basic architecture is Figure 2.1.

3. Methodology and Implementation. It might not be enough to merely take one or two elements into account when putting an accurate prediction model [15] into practice. Data on temperature, humidity, rainfall, and other variables are gathered and examined in Table 3.1.

The prediction model will be fed the results of this investigation. In this model using pandas [16] the yield_data and weather_data is imported and data is cleaned by detecting and removing the null values from the data sets. Removal of data Anomaly is carried out with the identification of the outliers using box plotting techniques and removing them. Forming a relationship between "weather_data and yield_data" using "district, year and crop_season" columns. Data Merging is carried by Creating 3 sub datasets from weather data grouped by "year", Taking average of Rainfall of each regiopn for same "crop_season", For every "year,district and crop_season" columns of "yield_data" , assign a rainfall value from the respective sub datasets that are chosen by year and from the chosen dataset "rainfall" data is selected based on "district and crop_season"

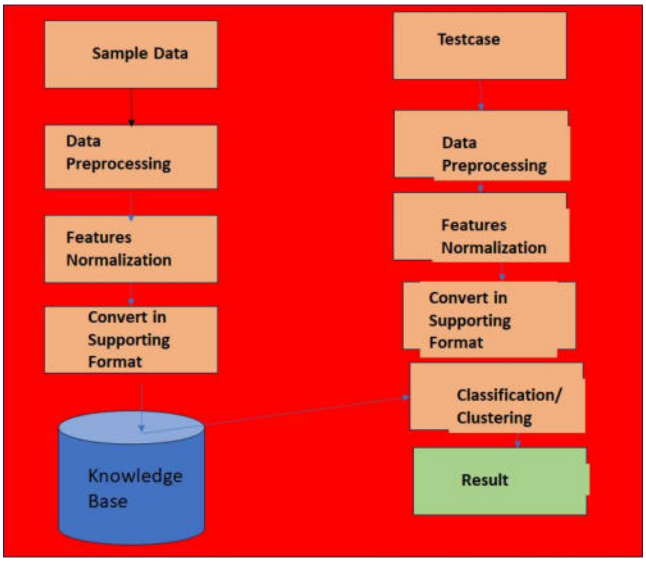

Fig. 3.1: Data Flow Diagram

columns and the new dataset created is then saved and used for further analysis. The Data Flow Diagram for the proposed model is illustrated in Figure 3.1.

3.1. KNN Algorithm. The KNN algorithm assumes that similar things exist in close proximity. The k-nearest neighbour (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm [17] that can be used to solve both classification and regression problems and the proposed system is implemented using Python [18] with feature scaling is demonstrated in Figure 3.2. The algorithm requires a labelled dataset with historical data of crop yields and corresponding input features includes weather conditions, soil properties, fertilizer usage, etc. The features should be numeric and normalized for better results. Predicting is carried out with a new set of input features for a specific crop, KNN searches for the K nearest neighbours from the training dataset based on a distance metric like Euclidean distance. It then predicts the crop yield based on the average or weighted average of the yields of those neighbours. KNN is relatively easy to implement, but it may be computationally expensive for large datasets.

Fitting KNN Classifier to the training set, after executing the output generated is as in Figure 3.3 with confusion matrix.

3.2. Naïve Bayes. Each feature in the Naive Bayes model [19] is presumed to be independent. In order to produce a good impression, everything must be the same. Based on this data, we can define something as a fact or a hypothesis: There is no association between any two features, according to our assumptions. The ”Hot” and ”Rainy” forecasts have little to do with humidity and wind conditions. As a result, we presume that the traits are distinct. The equal weighting of each attribute is the second factor (or importance). Temperature and humidity alone are insufficient for accurate forecasting and result of Naïve Base in Figure 3.4.

3.3. Random Forest Model. The Random Forest Machine Learning method [20] solves the problem of overfitting and improves the accuracy. The algorithm is as follows in Table 3.2 and implemented using Python.

4. Results and Discussions. The agricultural practises of organic farming practised in Telangana [21] were taken into account by the model as a case study, and they can work as a catalyst to boost crop output and management. Currently used methods of organic farming include preserving soil quality and promoting biological activity. indirect crop nutrient supply through the use of soil microbes. utilising pulses to measure the amount of nitrogen in the soil. To control weeds and pests, some organic techniques are utilised, such as crop rotation, natural predators, and organic fertilisers. Step-by-step gardening involves some in-depth conversation.

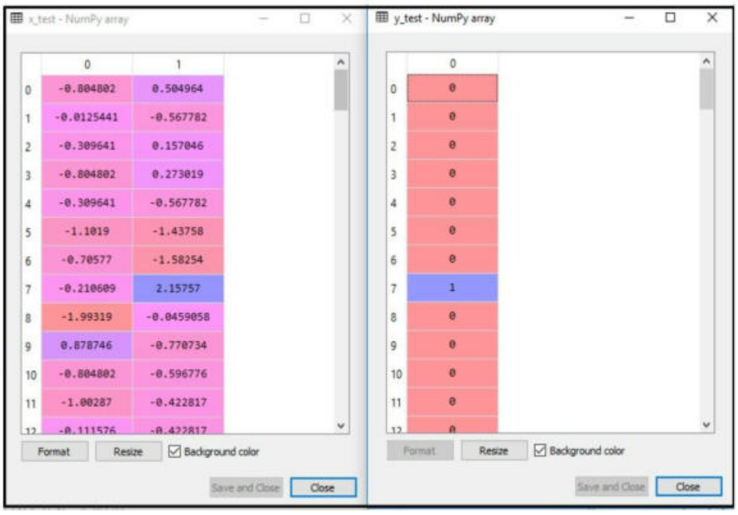

Fig. 3.2: Scaled Data

Fig. 3.3: KNN Output with Confusion Matrix

The first step is cropping selection, which entails selecting the appropriate crop for the season. Analysis of data is carried out based on the seasonal yields of major districts of Telangana state as in Figure 4.1. Maximum production is seen Warangal-Rural District.

The overall production of major crops in two different cropping seasons kharif [22] and rabi [23], the highest yield observed for the crops of maize and the least yield is for the crop of Bengal gram. The analysis of seasonal yield over selected years proves that maximum seasonal yield can be seen in rabi season as in Figure 4.2. The proposed system provides accuracy of 91.50% with naïve Bayes, 93.4% with KNN and 99.32 with Random Forest model.The highest rainfall was observed and recorded in northern regions of the Telangana State, Jainad Mandal in Adilabad district and Figure 4.3 demonstrates the line plot of data predicted by model. Integrated farming techniques are extremely useful in the mitigation of negative impact of agriculture or livestock on environment.

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Fig. 3.4: Naïve Base Result

Fig. 4.1: Seasonal yields of Telangana Districts

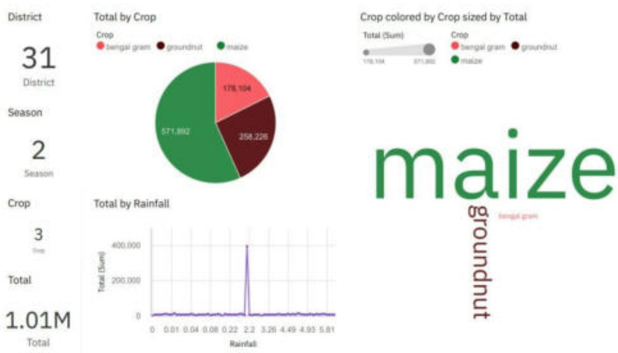

Fig. 4.2: Seasonal Yields of Crops

Accuracy of Naive Bayes not be as high as other more complex algorithms, but can provide decent precision

Table 3.2: Radom Forest Model - Python implementation and algorrithm

```
import pandas as pd

# Load the data from CSV
data = pd.read_csv("crop_yield_data.csv")

# Split the data into features (X) and target labels (y)
X = data.drop(columns=["Crop_Yield"])
y = data["Crop_Yield"]
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ranc
y_pred = rf_model.predict(X_test)

# Evaluate the model's performance
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

Algorithm

1. Import data (yield_data , weather_data)
2. Data Cleaning (Removing Null Values)
3. Anomaly detection (outliers using box plotting techniques)
4. Initialize K.
5. Identify Nearest Neighbour (weather_data and yield_data" , "district, year and crop_season").
6. Dataset Merging (year,district and crop_season, yield_data, rainfall, district and crop_season)
7. Data Transformation (label encoding)
8. Filter the data (Dist_id, Season, Crop, Rainfall, Season_yield)
9. Pick the first K entries from the Filter the data.
10. Prediction (x_train, y_train, x_test, y_test).
11. Evaluation (r2_score).

[49], especially when assumptions of conditional independence hold. However, it may struggle with precision if the features are strongly correlated [50]. Recall [51] of Naive Bayes can have good recall, particularly if it handles imbalanced classes well. It can effectively identify crops with low yields.

The F1 score of Naive Bayes will depend on both precision and recall, and it can achieve a reasonable balance between the two in many cases. Accuracy of Random Forest is resulting in high accuracy for crop yield prediction. Random Forest can achieve high precision, especially when the trees are well-optimized and the features are informative [52]. Random Forest tends to have good recall, as it combines multiple decision trees to capture different patterns in the data. Random Forest can achieve a high F1 score by balancing precision and recall, as it combines multiple decision trees with different strengths. The comparison of precision, recall and F1 score of three models are demonstrated by Figure 4.4.

5. Conclusions. The machine learning approaches analyse Meteorological data and provides accurate yield prediction to assist farmers. The model supports 99.32% accuracy with Random Forest and analyses Meteorological data of state Telangana. Areas need to be emphasized are increasing the focus for small farm economy utilizing market intelligence along with the State Governments involvement. The future work con-

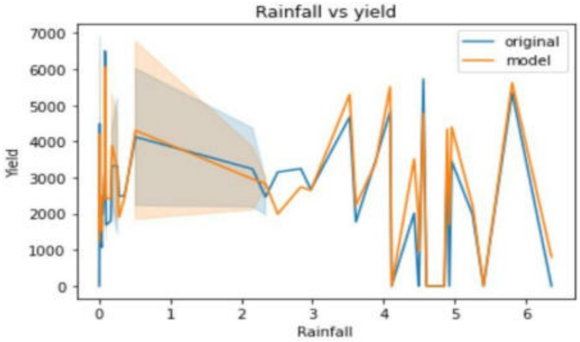

Fig. 4.3: Crop Yield Prediction

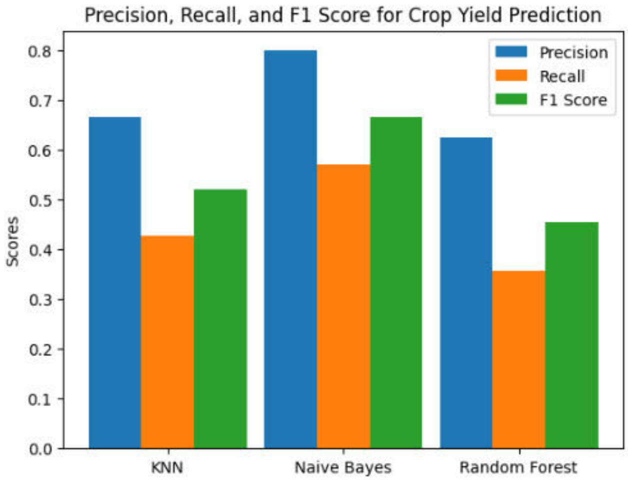

Fig. 4.4: Precision, Recall and F1 Score Comparison

centrates on artificial Supply chain management and Business intelligence strategies for marketing the crop of farmers.

REFERENCES

[1] Andrew Crane-Droesch, *Machine learning methods for crop yield prediction and climate change impact assessment in agriculture* , 2018.

[2] Kumar, Y. Jeevan Nagendra & Spandana, V. & Vaishnavi, V.S. & Devi, V.G.R.R, Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. 736-741. 10.1109/ICCES48766.2020.9137868,2020.

[3] Anna Chlingaryan, Salah Sukkarieh, Brett Whelan , Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture, Volume 151, Pages 61-69, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2018.05.012, 2018.

[4] Dr. G. Suresh, Dr. A. Senthil Kumar, Dr. S. Lekashri and Dr. R. Manikandan , Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming", International Journal of Modern Agriculture, Volume 10, No.1, ISSN: 2305-7246,2021.

[5] R. Seroul and S. Levy, *Aruvansh Nigam, Saksham Garg and Archit Agrawal; Parul Agrawal* , Crop Yield Prediction Using Machine Learning Algorithms", IEEE, 2019 Fifth International Conference on Image Information Processing (ICIIP), DOI: 10.1109/ICIIP47207.2019.8985951, 2019.

[6] D. Jayanarayana Reddy and Dr Rudra Kumar , Crop Yield Prediction using Machine Learning Algorithm", IEEE, Con-

ference: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), DOI: 10.1109/ICICCS51141.2021.9432236, 2021.

[7] F. Du, J. Zhang, J. Hu, and R. Fei , Discriminative multi-modal deep generative models,'' Knowl.-Based Syst., vol. 173, pp. 74–82, Jun, 2019.

[8] S. Zhang, W. Huang, and C. Zhang , Three-channel convolutional neural networks for vegetable leaf disease recognition,'' Cognit. Syst. Res., vol. 53, pp. 31–41, Jan, 2019.

[9] Elavarasan, D.; Vincent P M, D.R.; Srinivasan, K.; Chang, C.Y , A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling. Agriculture 2020, 10, 400. https://doi.org/10.3390/agriculture10090400

[10] Anna Chlingaryan, Salah Sukkarieh, Brett Whelan , Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture, Volume 151, 2018, Pages 61-69, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2018.05.012.

[11] Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D , Machine Learning in Agriculture: A Review. Sensors 2018, 18, 2674. https://doi.org/10.3390/s18082674.

[12] P. Priya, U. Muthaiah M. Balamurugan , Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[13] A. Nigam, S. Garg, A. Agrawal and P. Agrawal , Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019, pp. 125-130, doi: 10.1109/ICIIP47207.2019.8985951.

[14] A. Koirala, K. B. Walsh, and W. Z. McCarthy , Deep learning for realtime fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO,'' Precis. Agricult., vol. 20, no. 6, pp. 1107–1135, 2019.

[15] Dance Han Jeong, Jonathan P. Resop, Ethan E. Steward, Dennis J. Timlin, Kyo-Moon Shim, James S. Gerber, SooHyung Kim , Random forest for global regional crop yield predictions", 2016.

[16] Zhou, L., Pan, S., Wang, J., Vasilakos, A.V ,Machine learning on big data: Opportunities and challenges. Neurocomputing 237, 350–361 (2017).

[17] Kumar, S., Tejani, G.G., Pholdee, N., Bureerat, S , Performance enhancement of meta-heuristics through random mutation and simulated annealing-based selection for concurrent topology and sizing optimization of truss structures. Soft Computing 26(12), 5661–5683 (2022).

[18] S. Mishra, D. Mishra and G. H. Santra , Applications of machine learning techniques in agricultural crop production: a review paper", Indian J. Sci. Technol, vol. 9, no. 38, pp. 1-14, 2016.

[19] E. Manjula and S. Djodiltachoumy, A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, vol. 6, no. 4, pp. 2349-6363, 2017.

[20] Karan deep Kauri , Machine Learning: Applications in Indian Agriculture", International Journal of Advanced Research in Computer and Communication Engineering, April 2016.

[21] P. Saini and B. Nagpal , Efficient Crop Yield Prediction of Kharif Crop using Deep Neural Network," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 376-380, doi: 10.1109/CISES54857.2022.9844369

[22] Mulik, Sujata, Analysis of Crop Yield Prediction of Kharif & Rabi Jowar Crops Using Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 7. 79. 10.23956ijarcsse.v7i11.468, 2017.

Edited by: Vadivel Ayyasamy
Special issue on: Internet of Things (IoT) and Autonomous Unmanned Aerial Vehicle (AUAV) Technologies for Smart Agriculture Research and Practice
Received: Feb 1, 2024
Accepted: Apr 29, 2024