RESEARCH ARTICLE

PREDICTION OF CROP YIELD PRODUCTION SYSTEM: A MACHINE LEARNING APPROACH

Sandesh S. Waingade*, Nitin S. Chavan,

A. M. Chopade, L. G. Mulik and N. S. Kulkarni

Department of Computer Science,

Dr. Ghali College Gadhinglaj, Kolhapur, M.S., India

*Corresponding author E-mail: sandeshwaingade3999@gmail.com

DOI: https://doi.org/10.5281/zenodo.15173214

Abstract:

Agriculture plays a vital role in the economy of many countries, especially India, where the majority of the population depends on farming. However, farmers often face challenges in selecting suitable crops for cultivation due to changing climatic conditions, soil fertility, and availability of resources. This research focuses on predicting crop yield and identifying the most suitable crop based on soil and environmental features using machine learning techniques. The study uses features like Nitrogen (N), Phosphorus (P), Potassium (K), Rainfall, pH level, and Soil Type to predict crop production. A Linear Regression algorithm is applied to train the model and predict crop yield. The results of the model can assist farmers in making better agricultural decisions, ultimately increasing productivity and profit.

Keywords: Crop Yield Prediction, Machine Learning, Linear Regression, Agricultural Data, Soil Nutrients.

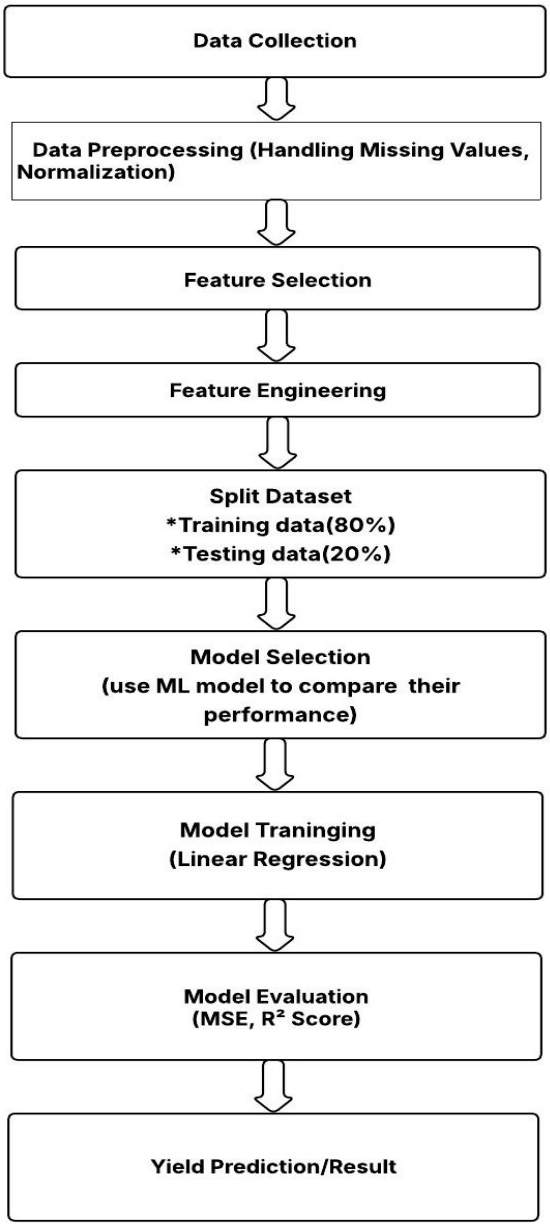
Introduction:

Agriculture is one of the most crucial sectors in India, contributing significantly to GDP and employment. However, farmers still rely on traditional knowledge for crop selection, which often leads to poor yield due to unsuitable crop choices for a particular region.

Crop yield prediction is essential for improving agricultural productivity and providing valuable insights to farmers for better crop planning. Machine learning, with its ability to analyse large datasets and extract patterns, offers a solution to predict crop yield based on multiple factors.

This research work aims to develop a Crop Yield Prediction system using Machine Learning techniques that consider essential soil nutrients and environmental factors to predict the most suitable crop and estimate its yield.

Flowchart:



Methodology:

1. Dataset Description

The dataset used in this study includes agricultural parameters such as nitrogen (N), phosphorus (P), potassium (K), rainfall, pH, and soil type. These features play a crucial role in determining crop yield. The data was collected from various agricultural regions to ensure diversity in soil and climatic conditions.

	Taluka_Name	Crop_Varieties	N_kg_ha	P_kg_ha	K_kg_ha	pH	Rainfall_mm	Temperature_C	Crop_yield_q_ha
0	Chandgad	Ghansal	120	50	57	7.3	1900	30	35.5
1	Chandgad	Ghansal	110	45	47	7.5	1800	31	35.5
2	Chandgad	Ghansal	115	48	57	7.8	1850	34	35.2
3	Chandgad	Ghansal	120	42	58	6.8	1750	29	35.6
4	Chandgad	Ghansal	100	43	69	6.9	1775	24	36.5

2. Data Preprocessing

To ensure the quality of data, preprocessing steps such as handling missing values, normalizing numerical features, and encoding categorical variables (soil type) were performed. Feature selection was applied to retain only the most significant attributes affecting crop yield.

3. Model Selection

Linear Regression was chosen as the predictive model for this study due to its simplicity and effectiveness in modeling continuous outcomes. The relationship between soil nutrients and crop yield is often linear, making Linear Regression an appropriate choice for this problem.

4. Training and Evaluation

The dataset was split into 80% training and 20% testing sets. The model was trained on the training dataset, and its performance was evaluated using Mean Squared Error (MSE) and Rsquared (R^2) score to assess prediction accuracy.

Accuracy for following Models:

```
from sklearn.linear_model import LinearRegression, Lasso, Ridge
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error, r2_score
```

```
models = {
    'ls':LinearRegression(),
    'lss':Lasso(),
    'rg':Ridge(),
    'Knr':KNeighborsRegressor(),
    'dtr':DecisionTreeRegressor()
}

for name, mod in models.items():
    mod.fit(X_train_dummy, y_train)
    y_pred = mod.predict(X_test_dummy)

    print(f"{name} MSE : {mean_absolute_error(y_test,y_pred)} score {r2_score(y_test,y_pred)}")
```

```
ls MSE : 0.755858196580611 score 0.9577700594127079
lss MSE : 3.2004299016294953 score 0.3168575527470098
rg MSE : 0.8083647786594862 score 0.9493107227304952
Knr MSE : 2.771 score 0.46470755869181724
dtr MSE : 0.9983333333333334 score 0.924636992807708
```

Results:

Model	Mean Absolute Error	R^2 Score	Accuracy (%)
Linear Regression	0.7558	0.9577	95.77%
Lasso	3.2004	0.3168	31.68%
Ridge	0.8083	0.9493	94.93%
KNeighborsRegressor	2.7710	0.4647	46.47%
Decision Tree	0.9983	0.9246	92.46%

The Linear Regression model demonstrated a strong correlation between soil nutrients and crop yield. The evaluation metrics indicated a low error rate and a high R^2 score, confirming the model's reliability. The results highlight the importance of soil nutrients in yield prediction and suggest that data-driven approaches can enhance agricultural productivity.

Conclusion:

This research presents a machine learning-based approach for crop yield prediction using Linear Regression. The model effectively utilizes soil nutrients and climatic data to provide accurate yield predictions. Future work may include exploring additional machine learning algorithms and integrating satellite data for further improvements in prediction accuracy.

References:

1. FAO (Food and Agriculture Organization). (2021). Agricultural Production and Productivity in India. Retrieved from [FAO website].
2. Kumar, V., & Singh, R. (2021). Predictive Analysis of Crop Yield Using Machine Learning Techniques. Journal of Agricultural Informatics, 12(2), 45-62.
3. Mishra, A., Kumar, P., & Singh, D. (2020). Impact of Climate Change on Crop Production in India. Indian Journal of Agricultural Sciences, 90(4), 1234-1242.
4. Sharma, P., & Gupta, N. (2022). Economic Impact of Data-Driven Decision Making in Agriculture. Journal of Economic Perspectives, 36(1), 75-92.
5. Government of India. (2023). Agricultural Statistics at a Glance. Retrieved from [Government of India website].