ISSN No:-2456-2165

Crop Yield Prediction Using Machine Learning

[1]Meka.Viswas
[2]Kedarisetti.Sri.Naga.Veera.Venkata.Swami.Subha.Surya
[3]Kathula.Ramu
[4]Medapati. Rama.Phaneendra Reddy-,
Swarnandhra College of Engineering &Technology.

Abstract:- Agriculture is an important part of the Indian economy and more than half of the country's population earns their living from agriculture. Agriculture is an important part of the growth of human civilization through the raising of domesticated animals that produce food that enables people to survive. Machine learning is used to predict crop yield based on parameters such as rainfall, yield and weather. In addition to being an important decision-making tool for crop yield prediction, machine learning also supports crop production and crop production-related decision-making. commonly used algorithms It is a neural network device. Weather, climate and other environmental factors pose a long-term threat to agriculture. Machine learning (ML) is important because it provides decision support tools for crop forecasting (CYP) that can help make decisions such as which crops to plant and how during the growing season. The main limitation of neural networks is to reduce the relative error and efficiency of crop yield prediction. The main objective of crop forecasting is to improve crop production and various models are used to achieve this goal. This research helps make agriculture more efficient by demonstrating machine learning's ability to predict crop yields with high levels of productivity. Design can be a decision support tool for farmers, enabling them to make informed decisions on crop management, resource allocation and risk mitigation, ultimately increasing agricultural sustainability and food security. Using the results of this study, farmers will be able to make informed decisions by determining the yield of their crops before planting on their farms.

Keywords:- *Crop_Yield_Prediction; Logistic_Regression; Naive Bayes; Random Forest; Dataset.*

I. INTRODUCTION

Machine learning (ML) technology is used in everything from predicting mobile phone usage to analyzing consumer behavior in supermarkets. Machine learning has been used in agriculture for a long time. Predicting crop yield is one of the most difficult problems in horticulture, and to date there are many models to be applied and evaluated. This question requires a lot of data because the harvest is affected by many factors such as soil, weather, environment, use of compost and many seeds. This shows that predicting crop yield is a result of the development process and not an explicit interaction. Potential models can now easily calculate actual outcomes; but better production forecasts are still needed. Regression techniques were used to make future predictions, while descriptive models were used to gain insight from the collected data and explain the results. Mechanical engineering research presents many challenges when trying to develop high-performance models. Choosing the right algorithm for the job is crucial; The algorithm and underlying platform need to be able to handle a lot of data.

In the Indian economy, the main occupation of people living in rural areas is agriculture. According to 2018 statistics, its population makes up more than half of the country, but its GDP accounts for only 17% of the country. Today's farmers grow products based on centuries of experience. Since the time of traditional methods, there have been examples of such over- or under-production where the actual rules for growing a crop are not observed. This is a major source of stress for Indian farmers. Farmer suicides are increasing due to low productivity, inbreeding and high production costs. All over the world, soil is being lost, water is polluted, and farmers' incomes are slowly decreasing. This problem can be solved by trying new crops, but farmers have difficulty predicting the results of a single crop. If they can predict which crop will yield the most, they will make more money and support farmers. Forecasting is one of the latest changes in agriculture and has a huge impact on the industry. It uses satellite imagery, comprehensive data on soil parameters, weather conditions and crop history to achieve optimum results. Therefore, we chose to use the most effective regression models, compare their accuracy, and see which model best predicted yield when some weather and soil parameters were not entered.

India grows crops throughout the year. Crops are classified according to growing season. There are two growing seasons in India: Kharif season and Rabi season. Kharif and Rabi growing season is from July to October and October to March. Thereafter crop and weather data were combined into a single file. Crop history information such as crop region, name, season, and other climatic factors affecting the crop are used as independent variables. The yield will be extraordinary.

ISSN No:-2456-2165

II. LITERATURE SURVEY

- Madhuri Shripathi Rao[1] said that agriculture is the main source of income for many developing countries. Today's agriculture is a path to farming success and continuous improvement of agricultural technology. It is important for farmers to respond to changes in the world, traders, consumers, etc. It becomes difficult to meet their expectations. Some of the problems faced by farmers are: (i) Tackling climate change due to soil erosion and industrial emissions (ii) Inadequate soil nutrients, shortage of essential nutrients such as potassium, nitrogen and phosphorus in vegetables slow down the harvest.
- A.M., Abouelghar[2] said that Pre-harvest forecasting of crops can prevent bad situations and help decision makers use reliable and accurate strategies on food safety. Remote sensing has many awards in the field of crop monitoring and yield prediction related to soil, climate and biophysical and biochemical changes. Different technologies are available for crop monitoring and yield prediction, including multispectral and hyperspectral data, radar and lidar imagery.
- Ms. Pratistha Mathur[3] said that Agriculture is an important part of the economy in India. Population growth is a problem for food security. Population growth leads to increased demand, which requires farmers to use the same farmland to produce more crops to increase capacity. Technology can help farmers increase crop yields by predicting crop yields. The main objective of this article is to estimate crop yield using area, yield, yield and irrigated area. Four machine learning methods decision tree, linear regression, lasso regression and ridge regression were used to predict the yield. Cross-validation method was used for validity, and absolute error, mean square error and root mean square error were used for validity. Decision trees outperform other machine learning methods.
- Sangeeta, S.G[4] said that Agriculture is an important source of energy, especially in developing countries such as India. The use of age records in agriculture can change the decision-making environment and farmers can earn higher profits. Agriculture, which supports about a fraction of India's population, accounts for only 14% of India's GDP. A plausible explanation behind this is that farmers do not have sufficient judgment regarding yield estimates. There is no basis for telling farmers about the plants they will grow. Machine learning techniques, rainfall, temperature, humidity, etc. It aims to predict the best crop in a particular region by analyzing various atmospheric factors such as.
- Narasimhamurthy, V[5] said that Rice is India's most important crop after wheat. India ranks second in the world after China and is generally considered a major producer of rice, accounting for 20% of the world's total production. The cultivated area in 20 states of India is approximately 40 million hectares. India continues to be the world's largest exporter with exports exceeding 100 million tons. Sustainability and efficiency of growing crops under favorable conditions. Developing better methods for predicting crop productivity in different climates can help farmers and other stakeholders make better decisions in agronomy and crop selection.

III. METHODOLOGY

This work aims to develop a regression algorithm using Google Collaboration Tools and Jupyter Notebook. The modeling steps are shown in the figure. The first step is to get the dataset from the Kaggle website; then before documentation, including maintenance and construction documentation; then perform data analysis (EDA); For some models, machine learning and random forest algorithms should be used. information. The table below shows some information about different laptops and their prices according to their features. Data from Kaggle.com

	State_Name	District_Name	Crop_Year	Season	Crop	Temperature	humidity	soil moisture	area	Production
0	Andhra Pradesh	EAST GODAVARI	2001	Kharif	Jowar	36.0	35.0	45.0	310.0	461.0
1	Andhra Pradesh	EAST GODAVARI	2001	Kharif	Lemon	37.0	40.0	46.0	988.0	3824.0
2	Andhra Pradesh	EAST GODAVARI	2001	Kharif	Maize	36.0	41.0	50.0	755.0	2500.0
3	Andhra Pradesh	EAST GODAVARI	2001	Kharif	Mango	37.0	42.0	55.0	15820.0	191027.0
4	Andhra Pradesh	EAST GODAVARI	2001	Kharif	Moong(Green Gram)	36.0	40.0	54.0	483.0	324.0

Fig 1: Sample dataset for crop yeild

ISSN No:-2456-2165

This data can cleaned and explorated using machine learning techniques and that mostly sutable forRandom forest algorithms . The data flow can be represented as follows.

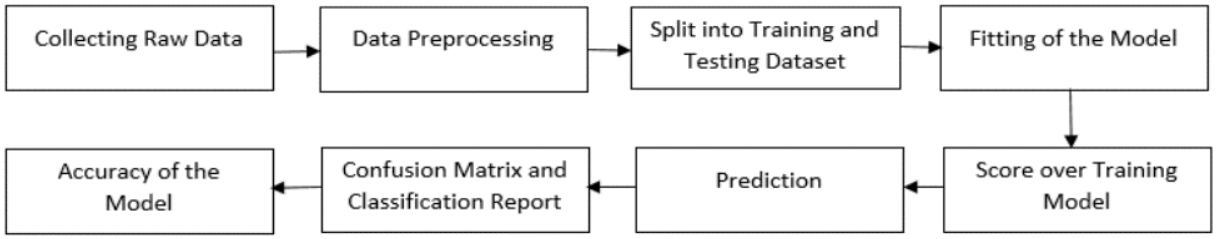

Fig 2 : data flow of model

Random Forest algorithm can be applied on the the dataset for the classification(To get the optimal price from the dataset). By using the random forest algorithm works 85% accurately.

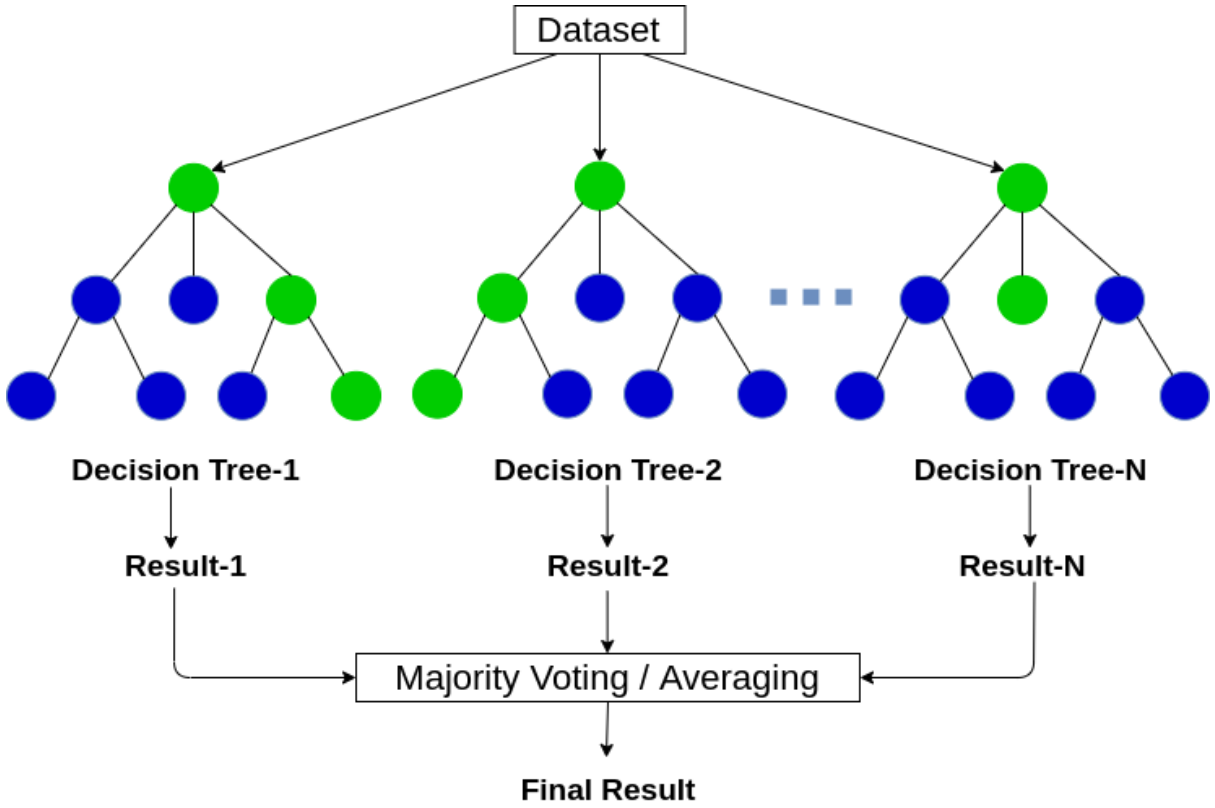

Fig 3: Random Forest

Random forests work in two stages; first, it creates a random forest by combining N decision trees, and second, it makes predictions for each tree created in the first stage.

We can explain the working process in the following steps and ways:
Step 1: Select the content K data from the training process.
Step 2: Create a decision tree with the selected data points (subset).
Step 3: Select the number N for the decision tree you want to create.
Step 4: Repeat steps 1 and 2.
Step 5: For new data, find the prediction for each decision tree and put the new data points in the class that won the most votes.

IV. APPLYING RANDOM FOREST

For this dataset we got the accuracy upto 85% by using random forest algorithm.

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error,r2_score,accuracy_score

X = data.drop('Production', axis=1)
y = data['Production']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create a Random Forest Regressor
reg = RandomForestRegressor(n_estimators=100, random_state=42)

# Train the model
reg.fit(X_train, y_train)

# Make predictions
y_pred = reg.predict(X_test)

# Check the Root Mean Squared Error (RMSE)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred))
print(f'Root Mean Squared Error: {rmse}')
```

```
Coefficient of determination: 0.85
Root Mean Squared Error: 16050251.891257439
```

Fig 4: working code of model.

V. CONCULUSON

The data we use in this analysis is provided by Kaggle and is based on facility data. In this analysis, we perform data cleaning, data analysis and data visualization. We concluded that product prediction can be made with 85% accuracy by using random forests and machine learning techniques for classification.

REFERENCES

[1]. Crop Forecasting Using Machine Learning Madhuri Shripathi Rao1, Arushi Singh1, N.V. Subba Reddy1 and Dinesh U Acharya1 1Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal, 576104, Udu, Karnataka, India Pi Region, 2022.

[2]. Ali, A.M., Abouelghar, M.A., Belal, A.-A. , Saleh, N., Younes, M., Selim, A., et al. (2022). Crop yield estimation using multiple remote sensing (see article). Sweetcorn. J. Remote Sensing Space Science. 25 Ib., 711-716. doi:10.1016/j.ejrs.2022.04.006

[3]. Kavita, Ms. Pratistha Mathur, Crop yield prediction using machine learning in India 2020 IEEE 5th International Conference on Computational Communications and Automation (ICCCA) (2020),

[4]. Sangeeta, S.G.: Design and use of predictive models used to produce crops. internationalization. J. education. machine. resources. 8(01) (2020)

[5]. Narasimhamurthy, V.: Rice yield prediction using random forest algorithm SML. internationalization. J. Res. application form. education. engineer. machine. 5, 1220–1225 (2017).