Crop Yield Prediction In Machine Learning Using RapidMiner Tool

Riya Varghese
Department of Computer Applications
Amal Jyothi College of Engineering,Kanjirappally,India
riyavarghese@mca.ajce.in

Meera Rose Mathew
Asst.Professor Department of Computer Applications
Amal Jyothi College of Engineering,Kanjirappally,India
meerarosemathew@amaljyothi.ac.in

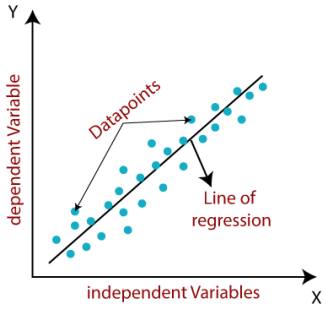
***Abstract*— Harvesting is a completely essential problem in agriculture. Machine learning is a critical subject of emerging research in crop yield evaluation. During the last decade, the climate has modified dramatically. Because of this, the farmers who were planting traditional crops are now facing problems. It is important that if farmers know the yield of the crop, they are planting in advance so that they can choose a crop that will suit their region. Analyzing various attributes such as season, crop type and area, will help farmers to know how much the crop will yield before harvest. The purpose of this paper is to help the farmer determine which crop will suit his region by predicting the yield of the crop. Machine learning is an important way to experience reality to overcome this problem. With the help of RapidMiner tool, Linear regression algorithm is used to train the model to get accurate predictions.**

Key words: Machine Learning, crop yield, RapidMiner, Linear Regression, Prediction

I. INTRODUCTION

About 58% of India's population's major source of livelihood depends on agriculture. In India, people cultivate several crops: Rice, Tobacco, Pulses, Wheat, Maize, Millets, Cotton, Jute, Sugarcane and many others. The livelihood of the farmers especially relies upon the harvest from the crop. Machine learning (ML) can be used for predicting the yield of the crop. Machine learning is an application for artificial intelligence (AI) which gives systems the capability to analyze automatically and improve from outdoor experience. There are distinct machine learning algorithms, such as linear regression, Decision Tree, Random Forest, Support Vector Machine, etc. we are using Linear regression algorithm for finding the predictions between the desired and the actual outcome of prediction. Linear regression is a statistical technique used for predictive evaluation. Linear regression algorithm shows a linear relationship between a dependent and one or more independent variables, which means it finds how the value of the structured variable is changing in line with the value of the independent variable. Linear regression version explains the relationship among a quantitative label and one or greater predictors(regular attributes) via fitting a linear equation to determined objects (with labels). Different tools like RapidMiner can be used for predicting crop harvest based on different parameters. By using machine learning tools, we can easily predict the crop yield in a particular state from the large data sets. Rapid Miner can be defined as a data mining tool which provides an integrated environment for machine learning, data mining, predictive analytics etc.

1.Linear Regression



The equation of Linear regression is given by,

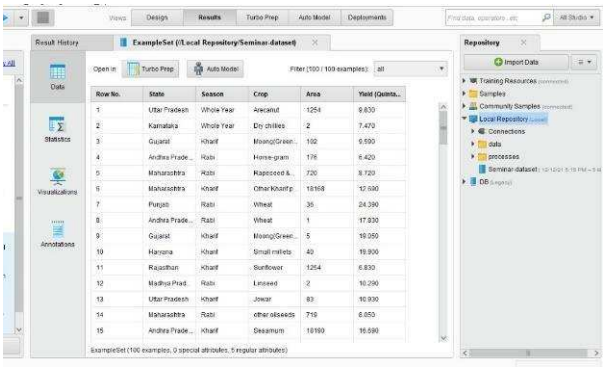
$$y = a_0 + a_1 x + \varepsilon$$

Y=DependentVariable
X=IndependentVariable
a0= intercept of the line
a1=Linear Regression Coefficient
ε = random error

II. METHODOLOGY

Data is a crucial part of crop yield prediction. In this research paper, data from different states of India are collected. The data about the different crops are gathered from the www.kaggle.com website. The data sets contain different states and their season, crop type, and area. By using these parameters, we can predict the yield of the crop.

DOI: 10.5281/zenodo.6369487
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam



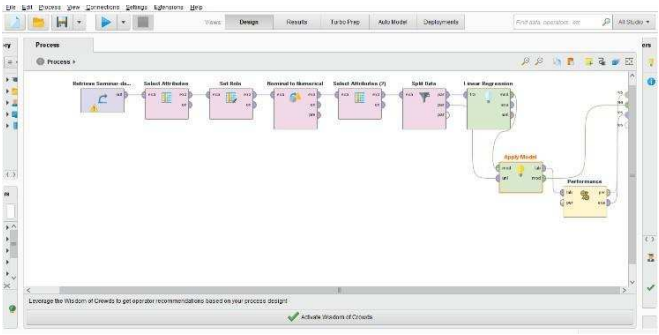
2.Dataset

A. Cropping Seasons in India

India has mainly three cropping seasons: Rabi, Kharif and Zaid. The crops which are grown between March and June are called summer crops, and the crops which grow over the year are called whole year crops. In this research paper, we have taken the season, crop and area which might be independent variables and we have taken the dependent variable as yield. This dataset is used for training the model through the use of the linear regression algorithm.

B. Implementation

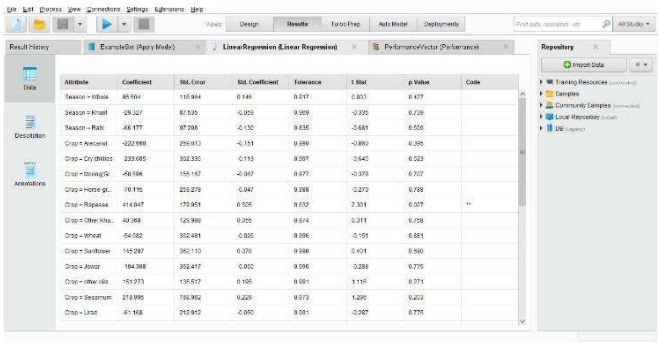
In this research paper, with the help of the RapidMiner tool, the prediction is done using the Linear regression algorithm. The dataset contains the yield of different crops from the different states based on season and area. In this predictive model, yield to be the structured variable and all other attributes to be the impartial variables. We have chosen the predictors of yield as season, crop, and area. In the regression version, all the predictors ought to be numerical. In the regression model, all of the predictors have to be numerical. Here the area is already numeric but the season and crop are not numeric. In order to convert qualitative characteristics to numerical characteristics, we use a nominal to numerical approach. In predictive modeling, we need to partition the dataset. In predictive modeling, we have to partition the dataset. Create a validation dataset (training the model) and test dataset for that we use split data. To measure the general overall performance of the version we use performance regression. After the design is finished, run the application to get the result.



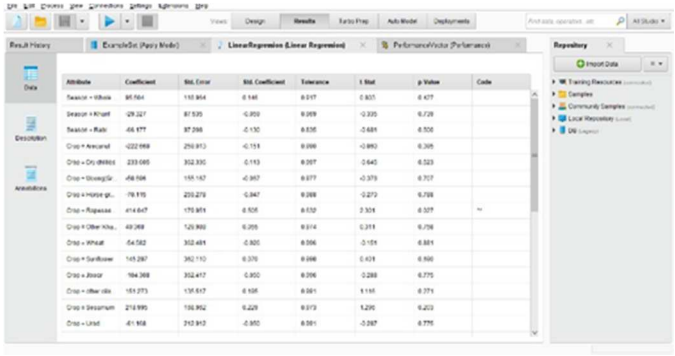
3.RapidMiner Tool used to design the model

C.Results

The aim of using the RapidMiner tool is to implement various machine learning algorithms such as linear regression algorithm to predict the yield of the crops. We get three results as the output. In the first result-ExamplSet (Apply Model) we can find the predicted value of yield by using the linear regression model. In the second result-LinearRegression (Linear Regression), we can find the coefficient and p-value of season, crop, area. In coefficient, the negative value means it has negative influence on yield and the positive value means it has positive influence on yield. P-value shows the significance. In the third result-performanceVector (performance) we find the root mean-square, it means on an average our regression model predicts the yield within a range +-327.939.



4.Expected value of yield with the aid of the usage of linear regression



5.coefficient and p-value of season, crop, area

6.Performance vector

DOI: 10.5281/zenodo.6369487
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

III. CONCLUSION

RapidMiner tool offers a graphical user interface to implement various machine learning algorithms. By using the linear regression model, we can find the predicted yield value of the crop and thus help the farmers to select suitable crops for their region. When you recognize the connection among the independent and based variables that have a linear relationship, this algorithm is exceptional to use because of its much less complexity in comparison to different algorithms. This process can be extended by taking different factors like soil parameters, use of fertilizers etc. that affects the yield of the crop.

IV. REFERENCES

1. https://www.irjet.net/archives/V6/i3/IRJET-V6I3281.pdf
2. https://www.irjet.net/archives/V7/i7/IRJET-V7I7421.pdf
3. https://indjst.org/download-article.php?Article_Unique_Id=INDJST9811&Full_Text_Pdf_Download=True#:~:text=Linear%20Regression%20(LR)%20is%20used,factors%20influencing%20the%20crop%20yield.&text=and%20their%20relationship%20with%20crop%20yield
4. https://www.jetir.org/papers/JETIR2102185.pdf
5. https://jpinfotech.org/crop-yield-prediction-and-efficient-use-of-fertilizers/

DOI: 10.5281/zenodo.6369487
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam