Big Data Analysis: Crop Yield Prediction

Sumukha K R
Information Science and Engineering,
RV College Of Engineering,
Bengaluru, India
sumukhakr.is17@rvce.edu.in

Utkarsh Singh
Information Science and Engineering,
RV College Of Engineering
Bengaluru, India
utkarshsingh.is17@rvce.edu.in

***Abstract*— India being an agricultural country ,its economy has a major dependency on agriculture as it contributes to 18% of the country's GDP and employs 50% of the population of India . Also in order to satisfy the needs of 138 crore people of the country, it is very crucial to have a good yield of agricultural crops. The crop yield relies on many factors like soil fertility or composition ,rainfall, precipitation, temperature and technological support.**

This paper presents an idea of a crop yield prediction system by analysing the agriculture dataset of the state Karnataka.The dataset contains the rainfall and temperature data for the district over years along with the Nitrogen(N), Phosphorous(P), Potassium(K) concentration of the soil . This data is used to construct a linear regression model to predict temperature and rainfall and a perceptron model which used the N,P,K values along with temperature and rainfall to predict crop yield. The accuracy of the system is checked using R-mean squared error(RMSE) and R-squared error.

Keywords— crop yield, machine learning, prediction, rainfall, temperature, karnataka, Linear regression, perceptron

I. Introduction

The Yield Prediction is an important issue in the field of agriculture as it depends on environmental factors. Farmers have considerations in knowing the yield of crops to plan the additional requirements if any to improvise the yield. Various parameters of weather, soil etc can be taken into consideration to analyse the crop yield so that the harvesting can be made better . In this paper, the yields of Rice and Wheat crops are predicted for some of the districts in Karnataka.

II. Motivation

Agriculture being the backbone of Indian economy demands utmost importance . The agricultural yield as can be seen every year is unstable because of some of the environmental factors which affect the economic conditions of the farmers and also the country . It also affects the import / export of goods and the cost of the product.

If technology can be used to support Agriculture to predict the factors which are affecting crop yield and then take measures to improve yield it would be an indispensable contribution. For Instance, if the rainfall predicted for a year is less, alternative measures can be taken for water supply to the fields to maintain moisture content of the soil. If the N, P, K values of the soil are less, fertilizers can be used to combat the difference. This motivates the study of crop yield prediction.

III. Literature Survey

Various research has been done for the crop yielding approach using data mining for the agriculture sector. Big data technology promises an efficient approach. But in comparison to traditional methods of handling data, handling large data has been the major issue. Niketa Gandhi,L J Armstrong[3] have achieved high accuracy rates using Sequential minimal optimization classifier with WEKA processing tool. They found high SD (standard deviation) while handling large datasets. Noronha , Divya, Shruthi [4] infer that data mining techniques provide a comprehensive evaluation of crop yield estimation. They have worked on the comparative analysis of data mining techniques. S. Pudumalar, E. Ramanujam [5] shows data mining for evaluating influences of the biotic and the abiotic components. Emphasize agriculturalists to identify precise crops based on the soil quality.

IV. Proposed Methodology

Crop Yield Prediction model using Perceptron and linear regression model.

1. Predicting next year rainfall using linear regression
2. Predicting next year temperature using linear regression
3. Training crop yield prediction model using perceptron
4. Inputs to predict yields :- N,P,K,Rainfall,Temp.
5. Output :- Yield (tons/hectare)

Dataset Description-

The dataset is downloaded from a Github repository and is divided into three parts- Temperature, Rainfall and Crop Yield.

The data in figure 1 is the sample data used to predict temperatures given the name of the district and year. The data in fig-2 is sample rainfall data which is used to predict the rainfall for a district and year. The data in fig-3 is sample crop yield data which has features such as rainfall,temp and N, P, K values for each crop grown in the districts of karnataka.

Figure 1- (temp for Bagalkot district)

District: Bagalkot

	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	1901	23.035	23.415	27.148	29.370	27.962	25.854	24.267	24.153	25.239	25.384	23.203	21.990
1	1902	22.898	24.153	27.870	29.130	29.325	26.665	25.146	25.309	24.501	25.396	22.636	22.397
2	1903	22.612	23.671	26.825	29.256	28.457	26.821	24.406	24.313	24.696	24.203	22.250	21.296
3	1904	22.120	23.569	26.586	29.041	28.463	25.561	24.271	24.882	24.752	25.087	22.118	21.535
4	1905	21.952	23.751	26.264	27.747	29.292	26.875	24.786	24.549	24.702	25.607	23.512	21.456

Figure 2- (Rainfall for Belgaum district)

	Year	Rainfall
0	1901	1901.616
1	1902	1735.258
2	1903	1778.020
3	1904	1562.044
4	1905	1039.236

Figure 3 - Crop(Bajra) yield for Bidar district

District: Bidar

	Crop	Year	Rainfall	Temperature	N	P	K	Yield
0	Bajra	1997	787.609	30.85	95.74	88.61	8.67	0.64
1	Bajra	1998	956.216	28.58	95.74	88.61	8.67	0.46
2	Bajra	1999	686.066	26.32	95.74	88.61	8.67	0.77
3	Bajra	2000	843.808	27.47	95.74	88.61	8.67	0.76
4	Bajra	2001	731.110	26.83	95.74	88.61	8.67	0.58

Necessary Packages

The packages mentioned below are used for reading, pre processing, training and testing the data-

a. Numpy
b. Pandas
c. Scipy
d. Matplotlib.pyplot
e. Scikit-learn

Data Preprocessing

The data for all the districts is first checked for existence and handling of any null values present in the data. The data is split into train and test data with 80% being training data and 20% as testing data, and for testing new values are also used. For example, if the model is trained on crop yield data for a particular district for a crop from 1998 to 2017 years, and then yield for that crop and district is predicted for years - 2018 and 2019.In the same way for any future years if one can gather data of the location of a particular crop then yield can be predicted.

Algorithms

Linear Regression model is used for predicting the temperatures for a given year and district by fitting the data into a quadratic polynomial. For predicting rainfall in a district for a given year, Linear regression model is used by fitting the data into a polynomial of degree 24 based on how well the model fits with the data. For predicting rainfall, perceptron(a single layer neural network) is used which associates weights and bias to the input data and predicts yield for a specific crop such as bajra,rice, jowar and groundnut etc given a district name, crop and year.

V. Results And Conclusion

TABLE-1: RMSE and R-Squared values while predicting rainfall for various districts

DISTRICT	**RMSE**	**R-SQUARED**
Bijapur	*0.13*	*0.098*
Chitradurga	*0.12*	*0.005*
Hasan	*0.11*	*0.005*
Tumkur	*0.12*	*0.00*
Belgaum	*0.11*	*0.02*
Bagalkot	*0.12*	*0.08*
Mysore	*0.11*	*0.01*
Bellary	*0.12*	*0.009*
Davangere	*0.12*	*0.02*
Raichur	*0.13*	*0.02*
Kolar	*0.12*	*0.02*
Kalburgi	*0.13*	*0.03*
Koppal	*0.12*	*0.01*
Gadag	*0.12*	*0.07*
Kodagu	*0.09*	*0.04*
Bidar	*0.13*	*0.01*
Chikmangaluru	*0.12*	*0.02*
Mandya	*0.12*	*0.00*
Shimoga	*0.11*	*0.04*
Dharwad	*0.11*	*0.12*
Chamrajnagar	*0.11*	*0.02*
Haveri	*0.11*	*0.05*

Rainfall and temperature prediction for a given year and district name was implemented successfully using a Linear Regression model. The RMSE or root mean square error and R-squared errors obtained while predicting rainfall for various districts are shown in the table-1, whereas the RMSE and R-Squared values obtained while predicting temperature for the given district are shown in table-2. In table-3, the yields predicted for Rice and Wheat crops in a few districts of Karnataka are shown. In future, the yields of crops-Rice and Wheat can be calculated for other districts of Karnataka and also developing a web application for it and make the user use it easily and help the user understand the yield of a particular crop for the place or the district in which the user lives.

Table-2: *RMSE and R-Squared values while predicting temperatures for various districts*

DISTRICT	RMSE	R-SQUARED
Bijapur	*0.026*	*0.112*
Chitradurga	*0.022*	*0.11*
Hasan	*0.019*	*0.147*
Tumkur	*0.022*	*0.114*
Belgaum	*0.023*	*0.160*
Bagalkot	*0.026*	*0.124*
Mysore	*0.019*	*0.159*
Bellary	*0.023*	*0.107*
Davangere	*0.021*	*0.138*
Raichur	*0.026*	*0.097*
Kolar	*0.022*	*0.094*
Kalburgi	*0.027*	*0.101*
Koppal	*0.024*	*0.112*
Gadag	*0.024*	*0.143*
Kodagu	*0.019*	*0.147*
Bidar	*0.029*	*0.101*
Chikmangaluru	*0.019*	*0.016*
Mandya	*0.020*	*0.131*
Shimoga	*0.019*	*0.195*
Dharwad	*0.022*	*0.180*
Chamrajnagar	*0.020*	*0.142*
Haveri	*0.019*	*0.147*

TABLE-3: Crop yield predicted for few districts of karnataka(tons/hectare)

District	**Rice** (Actual,Predicted)	**Wheat** (Actual,Predicted)
Bidar	*1.86, 1.67*	*1.34, 1.42*
Raichur	*2.78, 2.55*	*0.81, 0.90*
Gadag	*2.58, 2.78*	*0.54, 0.18*
Belgaum	*2.30, 2.56*	*1.51, 1.15*
Bagalkot	*2.76, 2.94*	*1.20, 1.45*

VI. ACKNOWLEDGEMENT

This paper is a report of the project which was done at the Department of ISE , RV College of Engineering® Bangalore.

First of all we would like to thank our supervisor and Head of the Department, B M Sagar ,for giving us the opportunity to work on this project and for all their encouragement and support. We would like to thank everyone else at RV College of Engineering® for all their help and for making my time at RV College of Engineering® a great and memorable experience.

VII. REFERENCES

[1] S Bhanumathi , M Vinneth , N Rohit , "Crop Yield Prediction and efficient use of fertilizers", International Conference on Communication and signal processing (ICCSP) , 2019

[2] Geetika , Rohit Bajaj , "Crop Yield prediction using data mining : an efficient data modelling approach " , International Journal of Engineering and Technology , 2018

[3] Niketa Gandhi , Leisa Armstrong, Owaiz Petkar, Amiya Kumar Tripathy , "Rice Crop yield prediction using support vector machines", In computer science and software engineering ,13th International Joint conference IEEE, 2016

[4] Perpetua Noronha , Divya J , Shruti B S, "Comparative study of Data Mining Techniques in crop yield prediction ", In vol 5 International Journal of Advanced research in Computer and communication Engineering(IJARCCE), 2016

[5] S Pudumalar , E Ramanujam , R Harine Rajashree, C Kavya , T Kruthika , J Nisha, "Crop Recommendation system for precision agriculture ", In 8th International conference on Advanced Computing (ICoAC) , 2016