Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Sugarcane Crop Yield Prediction Using Data Mining Application by Weka Machine Learning Tool.

Dr. Sujata Mulik [1], Dr. Ajit More [2], Dr. Madhuri Pant [3]

1Assistant Professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development Pune, India

2 Professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development Pune, India

3 Assistant Professor, Department of Computer Science, Faculty of Science and Technology, Vishwakarma University Pune, India

Abstract: - Data Mining concept is very useful for to discover important hidden patterns from large data. About 65percent of population in Sangli district is engaged in agriculture activities with sugarcane largely cultivated in irrigated area. For this study we have collected data sample from sangli district. In agriculture domain, crop yield prediction is very important and crucial. Agriculture domain problem that remains to be solved on the available past (historical) data.

In this research paper researcher has focused on the impact of weather/climate parameters as rainfall, temperature. On the sugarcane crop productivity. Data mining is very innovative research area in agriculture crop yield analysis. Ms-Excel is used for graphical presentation of crop data and weather data parameters. Weka is a machine learning tool specially we can used for prediction of result variable. We can find association between dependent and independent variables. This work aims at finding reliable data mining techniques to achieve a high accuracy result of yield prediction. In this research paper we noted actual and predicted crop productivity using linear regression and Smoreg(SVM).

Keywords*:* Agriculture, Data Mining, Yield Prediction, Sugarcane, Linear Regression, Smoreg,,Zero – R,Weka, Ms-Excel.

1] Introduction

Data mining and agriculture field are interrelated with each other means both are two sides of one coin. Agriculture is backbone of Indian economy. The agriculture field is primarily depending on climatic factors. Accurate yield prediction is very serious problem in the agriculture domain. Data mining process is to extract the useful information from a massive data set and transform it into understandable form for analysis. Data mining is discovering new patterns, so it is called as KDD.KDD stands for knowledge discovery in database. In KDD we execute various stages like data cleaning, integration, selection and transformation, data mining applications, pattern evaluation and finally user can represent knowledge in graphical form as per the data nature. In this paper we have taken sugarcane data and three weather parameters data for finding the predicted crop yield. This work aims at finding reliable data mining techniques to achieve a high accuracy result of yield prediction. In this research we are going to plan and analyse the actual and predicted crop productivity using classification and cluster analysis (Linear regression, k-means and SMOreg) algorithms of data mining using weka machine learning tool.

2] Summary of Data

In this paper researcher used data of sugarcane crop the years from 2010 to 2016 for Sangli district of Maharashtra in India. The evaluation is considered for only Sangli district in Maharashtra. The data are taken in six input variables. The variables are 'Year', 'Rainfall', 'Min_Temperature','Max_Temperature', 'Crop area', 'Crop productivity'. The attribute 'Year' specifies the year in which the data is available in Hectares. 'Rainfall' attribute specifies the average rainfall in the specified year in Millimetres. 'Temperature' attribute specifies the average temperature in the specified year in Degree Celsius. 'Area of Sowing' attribute specifies the total area sowed in the specified year for that region in Hectares. 'Crop Productivity (Yield)' specifies in Tons per hectare. This paper we focused study on dependent and independent variables. Rainfall, Temperature are independent variables & crop productivity dependent variables.

3] Overview on Tool

3.1] Ms-Excel: MS-Excel is a powerful spreadsheet that is easy to use and allows you to store, manipulate, analyse, and visualize data. Excel's graphing capabilities allows you to summarize your data enhancing your ability to organize and structure your data.

3.2] Weka: Weka is (Waikato Environment for Knowledge Analysis) a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is free software available under the GNU General Public license. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to this functionality. Weka is a collection of machine learning algorithms for solving real-world data mining problems. Data mining tool predict future trends and behaviour.

Weka will only use columns that statistically contribute to the accuracy of the model. Accuracy can be measured in R-squared. It will throw out data analysis and it can ignore the field that don't help in creating a good model that let us detect patterns, predict output and come up with conclusions backed by the data sample. Weka is a collection of machine learning algorithms for solving real-world data mining problems. Data mining tools predict future trends and behaviour. The algorithms can either be applied directly on the dataset or called from your own java code application.

Training: A data set is used to estimate or train a model

Test: A data set independent of the training data set that we use to fine tune the estimate of the model parameters.

Validation: The process of testing the model with a dataset is different from the training set. It checks whether performance of the model is improving or remains constant for the train model using validate data

4] Data mining and applications

4.1] Linear Regression: This is popular method for numerical data predictions. Linear regression model is used to fit a linear relationship between a dependent variable (outcome variable/response variable) and a set of predictors (independent /input/regressors/covariates variables). The correlation coefficient measures the robustness of the relationship between two variables. Pearson's correlation coefficient is one of the most used correlation coefficients and measures the linear relationship between two variables. The value of the correlation coefficient, denoted as r, ranges from -1 to +1, which gives the strength of the relationship and whether the relationship is negative or positive. When the value of r is greater than zero, it is a positive relationship; when the value is less than zero, it is a negative relationship. A value of zero indicates that there is no relationship between the two variables.

4.2] Zero-R: It is the simplest classification method which relies on the target and ignore all predictors. It is classifier simply predicts the majority category.

4.3] Smoreg: Smoreg implements the supports vector machine for regression. This algorithm is developed for numerical input variables. It gives us correlation coefficient value is 1 on sugarcane data sample and weather

parameters data values. So Smoreg is strongly associated data mining techniques with predicted result. Its predicted value is also good or accurate form as compared with the linear and zero-r crop predicted results.

5] Result Discussion and Analysis

The below table shown the crop data and weather data values. We have depicted the year wise crop productivity and rainfall and minimum and maximum temperature data in graphical form using ms-excel tool. Year 2012-2013 and 2013-2014 crop productivity is almost same. In 2015-2016 crop productivity is high average rainfall and temperature were good.

Table: 1 Year wise weather and crop data

1.Year	2.Area in ha	3.Productivity in tonnes/ha	4.Rainfall	5.Min_Temp	6.Max_Tem
2010-2011	763	103	22.57	17.58	30.27
2011-2012	705	102	9.55	22.35	30.15
2012-2013	607	116	21.02	22.07	31.08
2013-2014	626	119	16.55	20.98	32.27
2014-2015	745	109	22.56	20.96	33.07
2015-2016	840	96	19.9	21.39	32.27

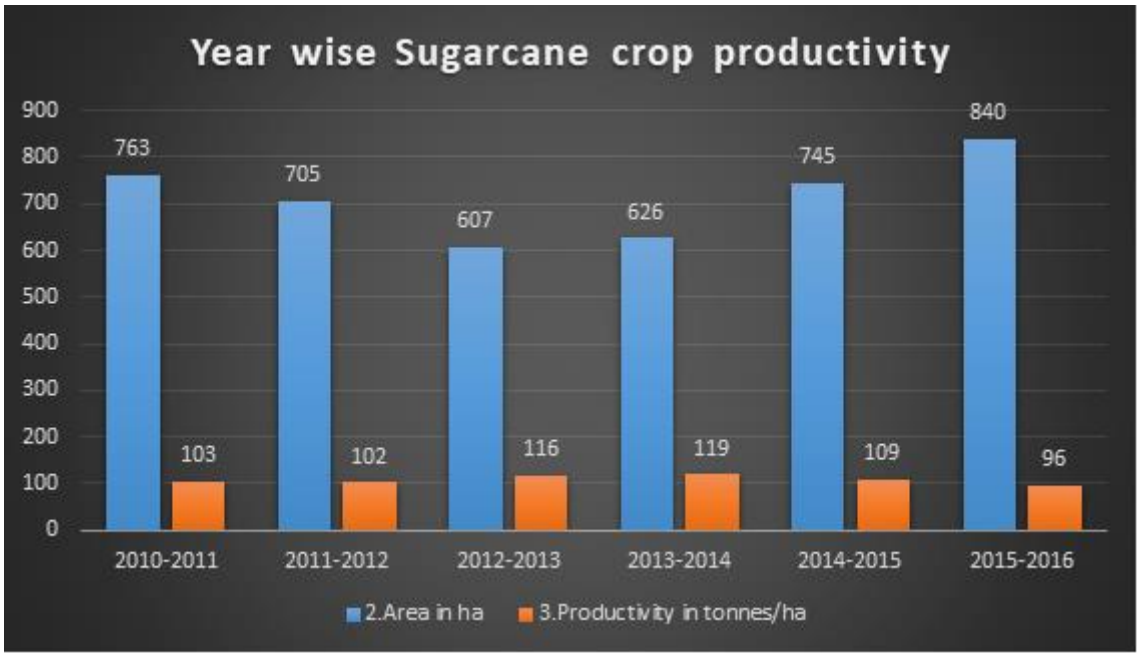


Figure 1: Year wise Sugarcane Crop Productivity

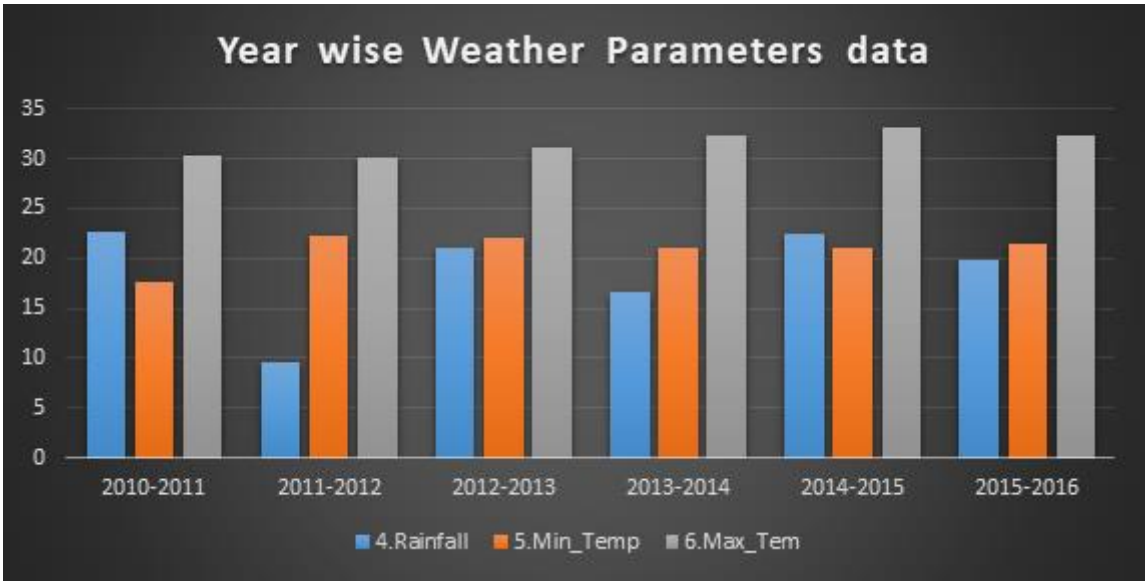


Figure 2: Graph of Year wise weather parameters data

Statistics Value	Crop Area in Ha	Productivity in ton/Ha	Rainfall	Min_Temp	Max_Temp
Minimum	607	96	9.55	17.58	30.15
Maximum	840	119	22.57	22.35	30.07
Mean	714.333	107.5	18.692	20.888	31.518
StdDev	87.763	8.826	5.001	1.717	1.197

Table 2: Statistical values of crop and weather parameters

	Linear Regression	Zero-R	Smoreg
Correlation coefficient	**0**	**0**	**1**
Mean absolute error	7.1667	7.1667	0.0223
Root mean squared error	8.0571	8.0571	0.0234
Relative absolute error	100%	100%	0.3116%
Root relative squared error	100%	100%	0.2909
Total Number of Instances	6	6	6

Table 3: LR, Zero-R, Smoreg algorithms results of crop and weather parameters

Linear	Rainfall	Min_Tem	Max_Temp
Correlation coefficient	**0**	**0**	**0.9995**
Mean absolute error	3.7611	1.1028	0.02
Root mean squared error	4.5651	1.5675	0.0346
Relative absolute error	100%	100%	1.964%
Root relative squared error	100%	100%	3.171%
Total Number of Instances	6	6	6

Table 4: Linear result of Weather parameters

Zero-R	Rainfall	Min_Tem	Max_Temp
Correlation coefficient	**0**	**0**	**0**
Mean absolute error	3.7611	1.1028	1.0183
Root mean squared error	4.5651	1.5675	1.0924
Relative absolute error	100%	100%	100%
Root relative squared error	100%	100%	100%
Total Number of Instances	6	6	6

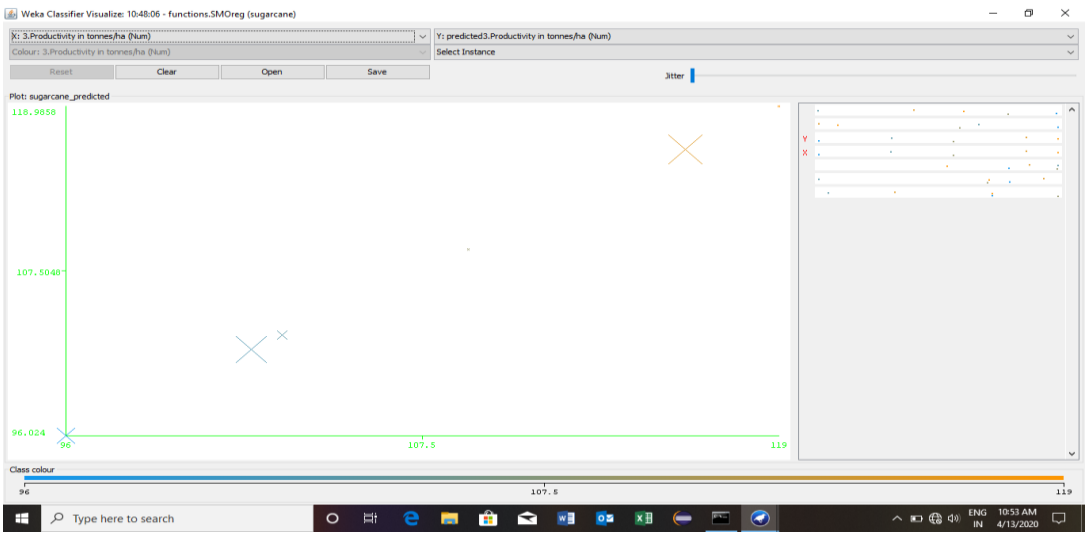
Table 5: Zero -R of Weather parameters

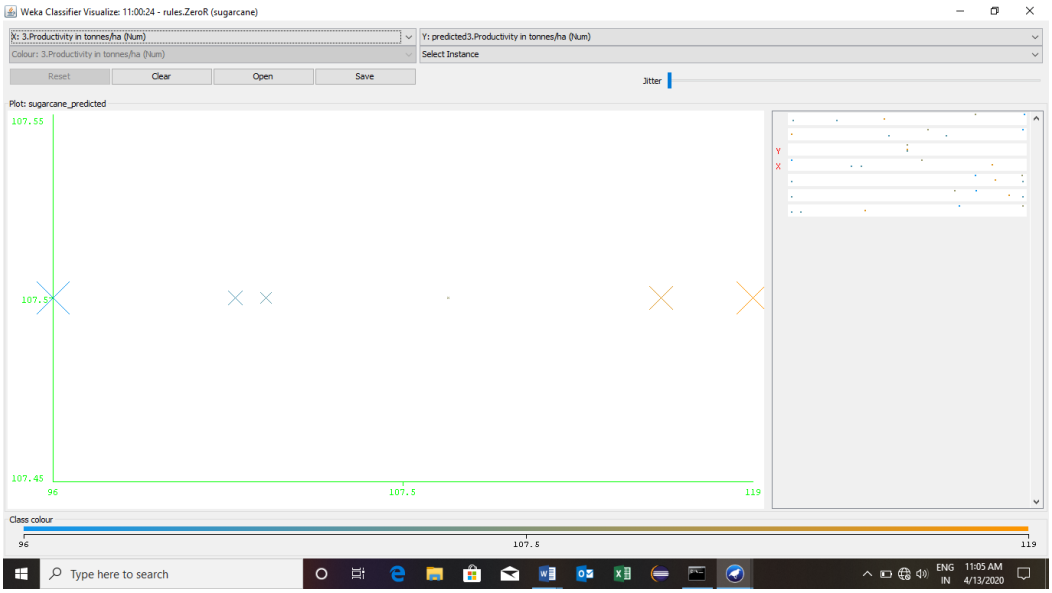
SmOreg	Rainfall	Min_Tem	Max_Temp
Correlation coefficient	**1**	**1**	**1**
Mean absolute error	0.0079	0.0058	0.0034
Root mean squared error	0.0099	0.006	0.0039
Relative absolute error	0.2106%	0.5217%	0.3355%
Root relative squared error	0.2166%	0.3846%	0.3603%
Total Number of Instances	6	6	6

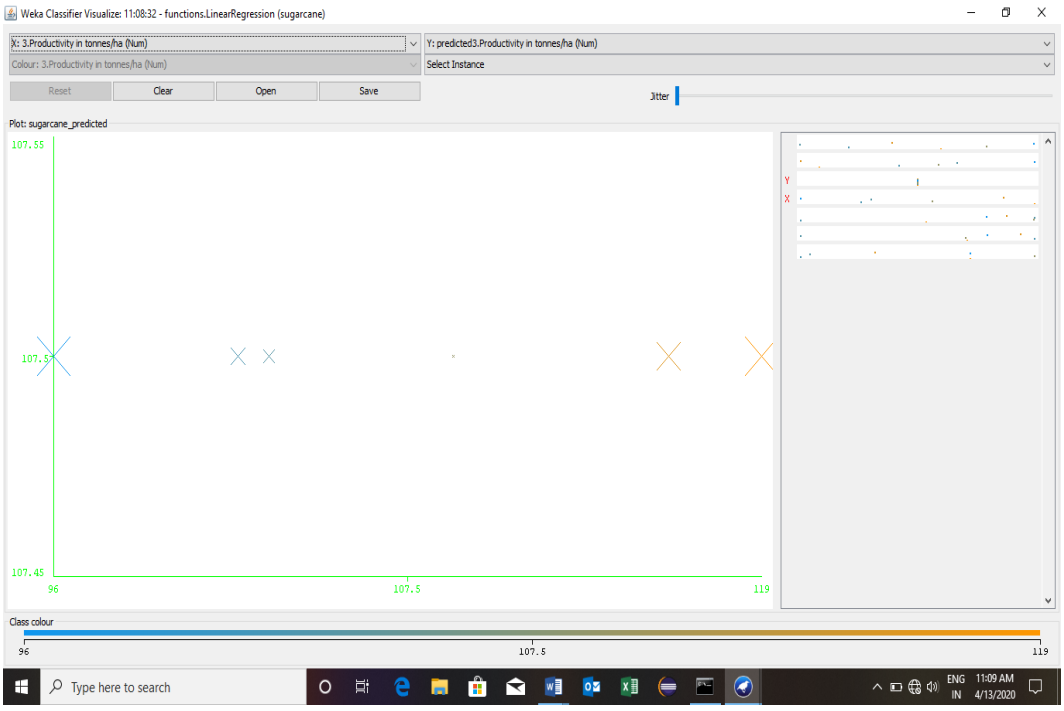
Table 5: Smoreg of Weather parameters

In the above analysis table, we have been seen that Smoreg data mining technique is very positively strong associated with the result as compared to linear and zero –r.In the above table we have shown the data analysis result of linear, zero-r and Smoreg technique. Also, we have presented the mean, maximum, minimum, and StdDev of crop area, productivity, rainfall and temperature.

The below we have shown the weka classifier result visualization.







Method			**Linear**	**Zero-R**	**Smoreg**
Year	Area in ha	Productivity in tone's/ha	Predicted Productivity	Predicted Productivity	Predicted Productivity
2010-2011	763	103	107.5	107.5	103.019
2011-2012	705	102	107.5	107.5	102.03
2012-2013	607	116	107.5	107.5	115.96
2013-2014	626	119	107.5	107.5	118.9

2014-2015	745	109	107.5	107.5	109
2015-2016	840	96	107.5	107.5	96.023

Table 6: Shown actual and predicted crop productivity using LR, Zero-R, and Smoreg.

In the above table we have stored the result of linear, zero-R, Smoreg sugarcane actual productivity and predicted productivity. Linear and Zero-R results are the same, but Smoreg productivity result is closest to sugarcane actual productivity. If the mean absolute error is less means it gives the closer prediction to actual value.

5] Conclusion

Data Mining is playing an important role in agriculture domain. We can use various kinds of data mining techniques for predicting the outcomes of variables. KDD is used to cleaning the bugs, errors, outliers, and inconsistency of data. In this paper we have used Ms-Excel and Weka tool for analysis the sugarcane data sample and weather parameters and to find the predicted crop productivity vs. actual productivity. We used three different data mining techniques in this paper. We found that Smoreg is strongly associated with the data sample. We found that some patterns of independent variables have an effect on crop yield productivity, this has been represented graphically in this paper.

References

[1] Ramesh1 , B Vishnu Vardhan2 , "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[2] Raj Kumar Tripathi , Nishtha Kesswani ," Clustering the Indian States on the Basis of Agriculture Produce of KHARIF and RABI Crops ",International Journal of Electronics Communication and Computer Technology (IJECCT) ISSN:2249-7838 Volume 2 Issue 2 (March 2012)

[3] Raorane A.A.1 Kulkarni R.V, "Review- Role of Data Mining in Agriculture", , International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013.

[4] Raorane A.A., Kulkarni R.V, "Data Mining: An effective tool for yield estimation in the agriculture sector", International journal of Emerging Trends & Technology in Computer Science,ISSN volume1,Issue2july – August2012.Rajshekhar Borate, 2Rahul Ombale, 3Sagar Ahire, 4Manoj Dhawade,

[5] Mrs. Prof. R. P. Karande ,"Applying Data Mining Techniques to PredictAnnual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Sftware Engineering ISSN 2394-7314 Vol. 3, Issue 1, pp: (34-37), Month: January-April 2016

[6] Mishra,S. ,Mishra D.,Santra ,G.H."Applications of Machine Learning Techniques in Agriculture Crop Production ". Indian Journal of science and Technology, 9(38) 2016

[7] Beulah "A Survey on Different Data Mining Techniques for Crop Yield Prediction" International Journal of Computer Sciences and Engineering, 7(1), 738-744.2019

[8] S R Jagtap" Feed forward neural network using back propagation for estimating efforts" International Journal of Innovative Knowledge Concepts, Vol.5 (11) 2017