

RICE YIELD PREDICTION USING NAIVE BAYES ALGORITHM

Kavya V K¹, Nishanth Hegde S², P Arihant Jain³, Pooja M Karnees⁴, Chandru A S⁵

1. Student, Information Science & Engineering, NIE Institute of Technology, Karnataka, India

2. Student, Information Science & Engineering, NIE Institute of Technology, Karnataka, India

3. Student, Information Science & Engineering, NIE Institute of Technology, Karnataka, India

4. Student, Information Science & Engineering, NIE Institute of Technology, Karnataka, India

5. Assistant professor, Information Science & Engineering, NIE Institute of Technology, Karnataka, India

ABSTRACT

Food is perhaps the most important aspect of life for, without it, humankind can't survive. Therefore, storing food becomes an absolute compulsion for countries all over the world especially in the developing ones. The critical aspect the governments of such countries want to fulfil is storing enough food for the long term, particularly at a time of natural disaster. Bangladesh is a low-lying region, hence the estimation of food is very demanding. In this paper, we attempt to forecast the yield of rice with astute analysis by using naïve bayes algorithm.

Key Words: rice prediction model, multiple linear regression, partial squares linear regression, datamining, agriculture.

1. INTRODUCTION

Crop forecasting or prediction is the art of predicting crop yields and production before the harvest actually takes place, typically a couple of months in advance. Crop forecasting relies on computer programs that describe the plant-environment interactions in quantitative terms. The soil testing program starts with the collection of a soil sample from a field. The first basic principle of soil testing is that a field can be sampled in such a way that chemical analysis of the soil sample will accurately reflect the field's true nutrient status.

1.1 Sub Heading 1

Increase in storage capability means that the repertoire of data is in an ever-burgeoning state. Storing huge amounts of data has allowed us to find patterns and create models that predict the future as accurately as possible. This research does precisely since its objective is to predict the yield of rice in the regions of Bangladesh during the aforementioned seasons using Multiple Linear Regression Ada Boost (Adaptive Boosting), Support Vector Machine Regression and a Modified Nonlinear Regression (MNR) equation and then comparing the modified equation with the other three methods to check its accuracy.

A lot of research has been done with the use of Data Mining, and the field of agriculture wasn't left out. A research was conducted in New Zealand in order to institute a pattern between the temperature difference and a grapevine yard. The attributes that had similar values are successfully clustered with the use of Neural Networks. Later, a chi square test was carried out to test the significance among the interrelated components. At the conclusion of the study, the results opined that the variance in temperature coupled with the likes of humidity, wind-speed and precipitation played an important character to determine the yield and quality of wine. important character to determine the yield and quality of wine.

The authors used Data Mining techniques—classification and clustering—and Neural Networks to picture the applications of Data Mining in Agricultural Yield. While selecting Rainfall as the dependent attribute and Year, Area of sowing and Production as the independent attributes, the authors used K-Means Algorithm and Multiple Linear Regression to make their prediction. With an accuracy of 98%, they concluded that MLR technique was more precise than K-Means Algorithm, which had a success rate of 96%. In this paper, the authors amalgamated Data Mining and Neural Network to predict the yield of wheat on a field in Germany, where they used factors such as vegetation, electric conductivity and fertilizer to make their predicted. The authors concluded that more data results to more accuracy. The authors made use of three attributes:rice yield, rainfall and yearly temperature—and used regression to find out the relation among them. In the end, it was concluded that temperature and rainfall does indeed have a say on the amount of rice yielded per season covers. Similar work was done in another study , as the

authors used Decision Support System for Agro Technology (DSSAT) to prove that temperature played an inverse role in the yield of rice.

2. BLOCK DIAGRAM

Multiple Linear Regression (MLR)-Linear regression is a method to obtain a relationship model between two variables, a dependent variable, Y with an independent variable, X. Multiple Linear Regression is almost the same thing, with the only difference being that in multiple linear regression, there can be a multiple number of independent variables. In this study, one of the techniques we used Multiple Linear Regression to predict the future yield. The dependent variable in this case is, obviously, the yield since it is the target variable. The rest of the attributes—area, rainfall, humidity and temperature—are the independent variables that are used in this research.

Adaptive Boosting (AdaBoost)-Adaptive Boosting is an ensemble language that does what its name suggests: it adapts with accordance to the data in order to build a sequence of complex predictors from simple predictors. The method it uses to generate such a pattern is by evaluating errors from the initial examples to then focus on getting these examples right and then combine the whole set with the use of adjusted weighted combinations to generate a complex classifier from several learning algorithms.

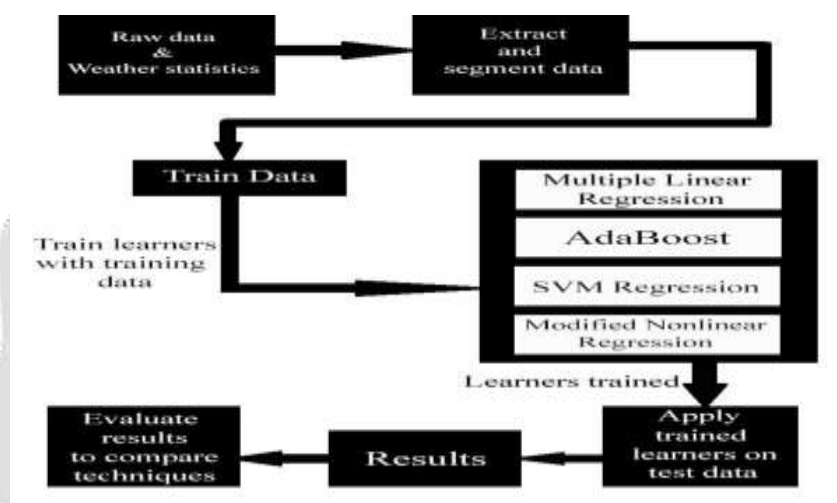


Fig -1: Architecture of RYPS(rice yield prediction system)

Support Vector Machine Regression-Support Vector Machine follows the concept of hyperplanes. In simple words, a line is drawn between the observations in a graph in such a way that it has the maximum possible distance from the sets of data points as it cuts through between them. That line is the hyperplane and the equation that defines the line is called a hyperplane equation. Unlike the other methods that have been listed thus far, SVM uses kernels in their equation to find out the prediction. The distance between a point and the hyperplane is the error and a point is only accepted if the error is smaller than a predefined accepted value of ϵ . The error values that are within the ϵ range are treated as if they are 0, i.e. as if they fall on the hyperplane itself.

Raw Data: The data collected from various departments regarding the crop details of previous years.

2.1 DEMO POINTS

The data used in this research is collected from the Agriculture Yearbook report generated by the Bangladesh Bureau of Statistics. It is a comprehensive report that consists of several attributes. However, some necessary data were missing and hence, WEKA averaging tool is used to find them out. In this study, the training data set is used to train the learner so that it could predict the yield from the test set. For the training set, data from 2011 -12, 2012 -13 and 2013-14 are used with Yield as dependent variable and Area, Rainfall Humidity and Temperature as the independent variables. Finally, the test data consisted only of the aforementioned independent variables for the year 2014-15 in order to predict the Yield from the same year. The predicted yields are then cross-validated with the actual yield from 2014-15 and from there, the RMSE, MSE, MAE and R-square values are calculated.

3. SYSTEM REQUIREMENTS

We need Hardware and Software requirements for this project.

3.1 HARDWARE REQUIREMENTS

- 2GB RAM
- 40GB Hard Disk
- Intel Pentium IV or advanced

3.2 SOFTWARE REQUIREMENTS

- Operating System : Windows 7 & above.
- Language : C#
- Design Tool : Visual Studio 2010.
- Back End : SQL server.

ADVANTAGES

Proposed system can be used in agriculture department to know the crop yield based on location and season wise. Proposed system can be used by farmers to know the rice yield by inputting data such as “region”, “humidity”, “area”, “temp” and “rainfall”. Proposed system can be used by government to know rice yield and hence properly maintaining the rice storage.

CONCLUSION AND FUTURE WORK

After interpreting the results of this study, it could be concluded that using our modified Nonlinear Regression equations works better than the other three predefined models, except while predicting Aman Rice, where it is marginally beaten by SVM Regression. For Aus and Boro Rice, however, our modified Nonlinear Regression equation not only yields better RMSE, MSE and MAE values, it also has the highest R-square value, which proves that the MNR equation is the best fit for this study. Furthermore, this study also proves that weather conditions play a very vital role in predicting the yield of rice. Indeed, with an accurate set of weather statistics, one can make an educated guess regarding the yield. This not only allows for future preparations in trade, but also allows for loss minimization. Data Mining can be used vigorously to help farmers yield the best possible amount of yield. The future work that one can carry out as an extension of this model is to use various predictions. Also, more statistics means that the accuracy of the results would increase. Hence, knowing the levels of osmosis, pesticides used and the amount of crop damaged in every season would further enhance the precision of the prediction.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible. So I would like to acknowledge all those whose guidance and encouragement served as a beacon light and crowned the efforts with success. I would like to sincerely thank the project guide, Chandru A S, Assistant Prof, Dept. of Information Science and Engineering, NIEIT for providing relevant information, valuable guidance and encouragement to complete this project.

REFERENCES

[1] S. Shanmuganathan, P. Sallis, and A. Narayanan, “Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine, Wine Quality and Perennial Crop Yield”, IEEE Second

International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), pp 90-95, 2010.

[2] D. Ramesh and B. V. Vardhan, "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.

[3] G. Ruß, R. Kruse, M. Schneider, and P. Wagner, "Data Mining with Neural Networks for Wheat Yield Prediction", P. Perner (Ed.): ICDM 2008, LNAI 5077, pp. 47–56, 2008.

[4] Md. A. R. Sarker, K. Alam and J. Gow, "Exploring the relationship between climate change and rice yield in Bangladesh", Elsevier Agricultural Systems vol 112, pp 11-16, October, 2012.

J. K. Basak, "Climate Change Impacts on Rice Production in Bangladesh: Results from a Model".

