

DATAMINING-A TOOL TO IDENTIFY CROP PERFORMANCE TRAITS FOR RICE VARIETIES OF WARANGAL REGION

VelidePhani Kumar^{1*} and Lakshmi Velide²

^{1*}Principal Consultant-Tele 9 Technologies, Hyderabad, Andhra Pradesh, India.

²AsstProfessor, Department of Biotechnology, Gokaraju Rangaraju Institute of Engineering and
Technology, Kukatpally, Hyderabad, India

Corresponding Author: +91-9704044407 Email: Phani.velide@hotmail.com

ABSTRACT: Crop related information is provided to the farmers by Government bodies and Private organisations. Farmers use this information to make farming profitable by selecting suitable varieties for that region. The information is not common to every farming station. Data mining techniques applied for provided crop research data customises it to each farming station. Present studies has been carried out on application of data mining in selection of suitable rice variety to Warangal region by obtaining the data of 20 years. The eight rice varieties recommended by rice research station, focussing on six traits were considered for the present study. Results show that Ramappa, JGL 384, Swarna, Samba masuri are high yielding varieties in Warangal region. Kavya, MTU 1010 and WGL14 have high harvested grain weight whereas Sureka, Samba Masuri and Ramappa have least. Kavya and WGL14 showed high straw strength, panicle length and carbohydrate content. This analysis helps the farmer to select particular combination of traits to identify high performing varieties for specific location.

Key words: Datamining, Rice varieties, Traits, Warangal region.

INTRODUCTION

When farmer intends to adopt new varieties, reliable information from seed marketer regarding the variety fastens the decision making process. Often shifting to new crop variety will result either in increase the yield or decrease. The seed marketer provides all the information to sell seed while farmer tries to know how true the information is. In this context the decision making becomes difficult to farmer as large number of potential varieties for farming are available. The Rice Variety Guide from Agriculture research stations provides key information and recommended varieties for specific region. The issue for the farmer is to integrate the information into a format that can then be applied to farming location, practices and decision making. This process of decision making often is where the farmer may seeks expert advice which can be furnished by computer based tools. Several studies have described the benefits to farmers in using such computer decision aid tools (Donnelly *et al.*, 2002 and Salmon *et al.*, 2007). Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used for industrial, commercial and scientific purposes. As such the process of data mining involves sorting through large amounts of data and discovering patterns in the data (Witten and Frank, 2005). Data mining is a part of OLAP (On line analytical processing) with application such as prediction in agriculture and so quickly answers multi dimensional queries. It provides an opportunity of viewing agriculture data from different points of view to better understand what that data means. Data mining has been used extensively for analysis of crop performance traits (Latikasharma and Nitu Mehta, 2012). Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and other analysis methods (Cunningham and Holmes, 1999). The recent advances in data mining technologies are successfully applied to the management of natural resources also. Several studies on cotton crops in Pakistan have been reported using data mining technologies by Abdullah and Ansari, (2005), Ekasingh *et al.*, (2005) and Holmes *et al.*, (1998). Armstrong *et al.*, (2007) have reported that data mining empowers farmers in selection of site specific crop varieties. The present research determined whether data mining techniques can be used to identify key traits of rice varieties and recognise fit variety for Warangal district.

HISTORY

Always new crop variety information can be obtained by research trails which later can be compared with the available one. The research trials always carried under regional farmer practices. The data will be combined and averages for each trail is measured.

The environmental and geographic condition will affect the values at each trial location. Farmers always doubts which factors effects the crop variety and on which trait at a particular location. Does any change in the practicing site will show impact on required traits. Data mining address these concerns by converting generalised data of multiple locations to specific data.

MATERIALS AND METHODS

Data (20 years) for this study was taken from Rice knowledge management portal (RKMP), Directorate of rice research, Rajendranagar, Hyerabad, Andhra Pradesh. The varieties selected for this work were Surekha, Swarna, Samba masuri, Kavya, Ramappa, MTU 10101,WGL14,JGL 384.These 8 varieties have been recommended by research station since 20 years for 150government owned trial locations. The traits (Indicators) selected for study were Crop yield (Y), Carbohydrate content (CH), Total straw, Straw strength (SS), Harvested grain weight (HGW/100grains), Panicle length (PL).

Statistical Analysis

For combined analysis datamining tools were used from R (R Development Core Team 2012). R and A Sreml (2012) –R mixed multivariate model was used to derive the predictions on the data. The predicted values were analysed with principal component analyses.The predicted values from mixed models were put into data cube in Postgresql (ref:Postgresql 2012), Open office.org (ref:Openoffice.org2012) and database driver (Postgresql-sdbc-driver2012). Reporting and querying predictions were done by data cube.

RESULTS AND DISCUSSION

Table 1 explains how predictions of rice traits effects by average process across 150 trial locations of Warangal region.Results show that all trait average values were relatively uniform.Fig1explains the relationship between various traits and varieties recommended by research station. The closer the trait to the prescribed variety line the greater the correlation between variety and the trait. Fig 2 depicts that ID 5,8,3and 2 are high yielding varieties in Warangal region.ID 4,6 and 7 have high harvested grain weight whereas 1,3 and 5 have least. ID 4 and 7 showed high straw strength, panicle length and carbohydrate content. This analysis helps the farmer to select suitable variety for specific location based on their interested trait.Thus data mining enabled the farmer to check the available data and search for the required trait bearing variety.It is also envisaged that data mining helps the researchers to search for a variety adapted to specific environment and also helps to improve some traits in the available variety to get adapted to the new environment.

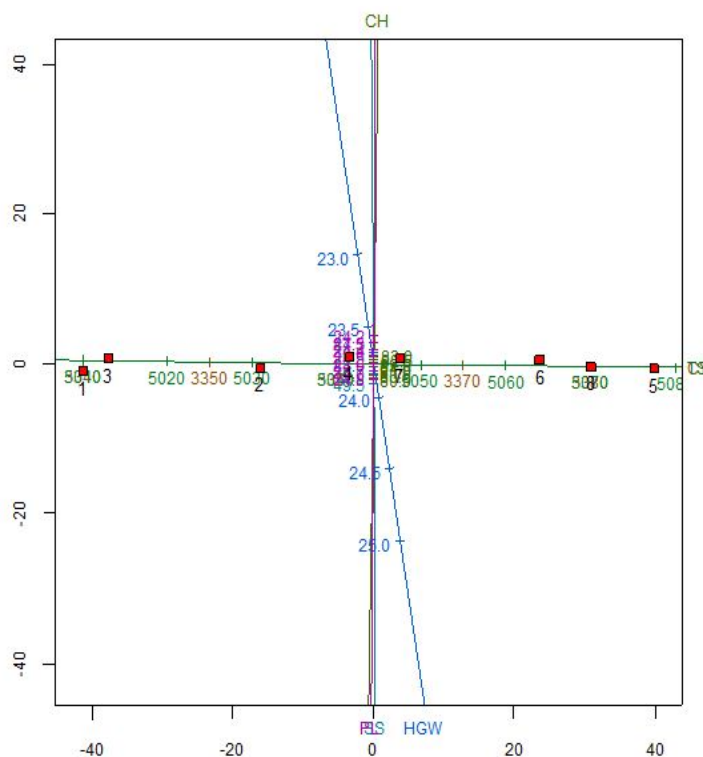


Figure-1: Multivariate biplot

Table:1 Predicted values for average locations and seasons

| Variety/ ID | Traits | | | | | |
|----------------|--------------------|-----------------------|---------------------|----------------------|----------------------------|---------------------|
| | Crop yield (Kg/ha) | Carbohydrate (g/100g) | Total Straw (Kg/ha) | Straw strength (Mpa) | Harvested grain weight (g) | Panicle length (Cm) |
| 1.Sureka | 3340 | 80.2 | 5010 | 48.3 | 23.5 | 24.5 |
| 2.Swarna | 3354 | 81.4 | 5031 | 49.2 | 23.8 | 24.8 |
| 3.Samba masuri | 3342 | 82.3 | 5013 | 48.7 | 23.2 | 24.8 |
| 4.Kavya | 3361 | 82.6 | 5041.5 | 47.6 | 24.2 | 24.6 |
| 5.Ramappa | 3385 | 81.8 | 5077.5 | 48.5 | 24.6 | 24.7 |
| 6.MTU 1010 | 3376 | 82.8 | 5064 | 48.7 | 23.8 | 24.3 |
| 7.WGL 14 | 3365 | 82.4 | 5047.5 | 47.9 | 23.6 | 24.2 |
| 8.JGL384 | 3380 | 82.2 | 5070 | 49.2 | 23.4 | 24.8 |

CONCLUSION

Farmers have large number of choices in selecting new varieties. Various tools helps them to identify the field performance data of different rice varieties. These tools can be obtained freely but some knowledge regarding application of tools for data interpretation is needed. Data mining helps the farmers in completion of the task easy. Thus in conclusion data mining helps to retrieve Warangal region rice variety data of twenty years and used to select suitable variety of that region with required traits.

ACKNOWLEDGEMENTS

The authors would like to thank directorate of rice research station, Hyderabad for providing the data.

REFERENCES

- Abdullah A and I A Ansari. (2005). Discovery of cropping regions due to global climatic changes using data mining. Paper presented at the CAIR Publications, Beijing China.
- ASReml.(2012) Available from <http://www.vsn-intl.com/products/asreml> (Accessed 28 December 2012).
- Cunningham S J and G Holmes. (1999). Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference.1999.
- Donnelly JR, MFreer, LSalmon, AD Moore, RJSimpson, H Dove and TP Bolger. (2002).Evolution of the GRAZPLAN decision support tools and adoption by the grazing industry in temperate Australia. *Agricultural Systems*.74 (1):115-139.
- Ekasingh BS, NGamsomsuke, RA Letcher and JM Spate. (2005). A data mining approach to simulating land use decisions: Modelling farmer's crop choice from farm level data for integrated water resource management. Paper presented at the Proceedings of the 2005 International Conference on Simulation and Modelling, 17-19 January, Bangkok, Thailand.
- Holmes G, L Hunt and RJ McQueen.(1998).User satisfaction with machine learning as a data analysis method in agricultural research. *New Zealand.J.Agri.Res.*41:577-584.
- Latikasharma and Nitu Mehta (2012).Data Mining Techniques: A Tool For Knowledge Management System In Agriculture. *Int.J.Sci.andTech.Res.*1(5):67-73.
- OpenOffice.org.(2012).Available from <http://www.openoffice.org> (Accessed 28 December 2012).
- Postgresql (2012).Available from <http://www.postgresql.org> (Accessed 28 December 2012).
- Postresql-sdbc-driver. (2012).Available <http://dba.openoffice.org/drivers/postgresql/index.html> (Accessed 28 December 2012).
- R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.ISBN 3-900051-07-0. Available from <http://www.r-project.org>.(Accessed 28 December 2012).

Salmon L and JR Donnelly.(2007).Using grazing systems models to evaluate business options for fattening dairy bulls in a region with a highly variable feed supply. *Animal Feed Sci.Tech.*5: 15-16

Witten I H and E Frank. (2007).“Weka 3: Data Mining Software in Java”. <http://www.cs.waikato.ac.nz/~ml/weka/> (Accessed 28 December 2012).