Research Paper Title- Analysis of Production of Tur crop in Akola District using Regression in WEKA Data mining tool and design a model.

Priyanka Thakare[1] & Dr. R.M.Patil [2]

[1] Asst.Prof. & Phd Regd. Student in Savitribai Phule Pune University.
[2] Professor, DYPIMR, Pune.

ABSTRACT: *In the current era, India ranks second worldwide in Agriculture. Maharashtra is one of the biggest, wealthiest and most developed states in India by current economic status. financial system of Maharashtra is basically depending on agriculture which have incredible amount of data which needs to be mine to improve the profitability of crops. Data mining in agriculture is a very recent research topic. It consists in the application of data mining techniques to agriculture. This research paper focuses on the analysis of tur as a kharip crop in Akola district using WEKA tool by applying linear regression analysis. For this the agro climatic data of Akola district collected from the websites of Indian Meteorological department and department of agriculture Maharashtra state. In this paper, researcher creates the regression model for production of tur crop in Akola district of Maharashtra using WEKA tool.*

Key Words: *Data mining,Agriculture, regression Analysis, WEKA tool.*

I. Introduction

Agriculture is back bone business in India. Agriculture plays a vital role in India's economy. Today, India ranks second worldwide in farm output. In Indian agriculture, large amount of data is available Data mining is used for mining data from databases and finding out meaningful patterns from the database with its great ability to dig out the hidden information. financial system of Maharashtra is highly depends on agriculture. It is the main occupation of the people. Both food crops and cash crops are grown in the state. The main food crops of Maharashtra are wheat, rice, jowar, bajra, and pulses. Cash crops include groundnut, cotton, sugarcane, turmeric, and tobacco.

But the scope of this research study mainly focus on Production of tur crop in Akola district of Maharashtra state. In Maharashtra basically the farmers are depends on traditional crops only and due to this they faces the economical problems due to less Production of traditional crops. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Data mining technique plays a vital role in the analysis of data. This research paper focuses on linear regression a data mining technique using WEKA tool. In the present research, this linear regression analysis is used to analyze the agro climatic data of Akola districts of Maharashtra which will predict the Production of tur crop .

II. Regression in WEKA – A Data mining tool

1. Data mining

Data mining is the transformation of large amounts of data into meaningful patterns and rules. Data mining is divided into two types: directed and undirected. In directed data mining, it will try to predict a particular data point. In undirected data mining, it will try to create groups of data, or find patterns in existing data. Additionally, the term data mining is all-encompassing, referring to dozens of techniques and procedures used to examine and transform data. The ultimate goal of data mining is to create a model, a model that can improve and interpret existing data and future data. In this paper, researcher creates the regression model for production of tur crop in Akola district of Maharashtra using WEKA tool.

2. WEKA Tool

Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. WEKA is the product of the University of Waikato (New Zealand) and was first implemented in its modern form in 1997. It uses the GNU General Public License (GPL). The software is written in the Java ™ language and contains a GUI for interacting with data files and producing visual results (think tables and curves). It also has a general API, so WEKA will embed, like any other library, in any applications to automated server-side data-mining tasks. Some features of data mining as shown in following figure1.

Figure.1-Features of WEKA tool

3. Regression

Regression is the easiest technique to use in data mining. Regression is a data mining function that predicts a number. Regression analysis is a form of predictive modeling technique which investigates the relationship between a dependent (target) and independent variable (s) (predictor). This technique is used for forecasting, time series modeling and finding the causal effect relationship between the variables. There are various types of regression techniques are present but linear regression is one of the most widely known modeling technique. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear. Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).

It is represented by an equation $Y=a+b*X + e$, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

In this study researcher use the linear regression technique to predict production(Y) of tur crop in Akola district based on area(x1), temperature(x2) and rainfall(x3) and design a model.

III. Objectives of the Study

1. To study the tur as a kharip crop in Akola district of Maharashtra.
2. To study the agro-climatic condition in Akola district of Maharashtra.
3. To find out the Production of tur as a kharip crop with the help of regression analysis a data mining technique which will helpful to improve the financial growth of farmer.
4.To bulid a regression model

IV. Research Methodology

The present research will consider the study area as Akola district of Maharashtra. This research will use regression analysis a data mining technique using WEKA tool to process and analyze various input agro climatic parameters such as rainfall, temperature, area and Production to produce output and designed a model.

V. Data Collection

This research is based on secondary data that has been collected by referring various research articles, books & different web-sites such as department of agriculture of Maharashtra and Indian meteorological department. Dataset consist of last 5 years data from 2013-2017 with following parameters namely: year, State-Maharashtra, District Akola, crop tur, season (kharif), area (in hectares), Production(in tones), average seasonal temperature (°C), average seasonal rainfall (mm).The data has been collected from Akola district of Maharashtra analyzed, compiled & then the outcome of all these are presented in this research article.

VI. A Review of Literature on Agriculture and Data mining techniques

A number of studies have been carried out on the data mining techniques for agricultural data sets. The literature review provided the background information and conceptual framework for the present research.

The research conducted by Mr. Abhishek B. Mankar, Mr. Mayur S. Burange (2014) mainly focused on the applications of Data mining techniques in agricultural field.

Raorane A.A., Kulkarni R.V. (2013) had described new data mining techniques and applied them to the various variables consisting in the database. Data mining techniques were applied in field of agriculture

as an objective methodology for pre-harvest crop forecasting. This research provided suitable forecast model(s) which had certain merits over the traditional forecasting method.

D Ramesh , B Vishnu Vardhan has explained (2013) data mining techniques such as K-Means approach and Bi-clustering to estimate crop yield analysis with existing data. Also explained there uses and benefits in agriculture sector.

S.S.Baskar, L.Arockiam, S.Charles (2013) described an analysis of the soil data using different data mining algorithms and prediction techniques.

The research done by Shweta Taneja, Rashmi Arora, Savneet Kaur, (2012) mainly used data mining technique in the field of soils. The main goals of the research were to find out the useful relationships among different types of soils.

J. Rajendra Prasad, P. Ravi Prakash, S. Sai Kumar, M. Sundara Babu K. Swarupa Rani (2012) proposed a data mining framework for the identification of agricultural production areas in Andhra Pradesh. This study is specially used for crop prediction.

S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh (2011) explained a review on application of data mining techniques in the field of agriculture.

The research and comparative study conducted by P. Bhargavi, Dr. S. Jyothi (2011).This research has proposed an analysis of the soil data in agriculture using different algorithms and prediction technique.

The research proposed by Ramesh Vamanan ,K.Ramar (2011) explained the data mining techniques can be applied effectively in agriculture filed to compare large amount of soil data

Dr. D. Ashok Kumar, N. Kannathasan done the research on (2011) various data mining and pattern recognition techniques for soil data.This survey suggested different data mining techniques used in agriculture soil science to extract new knowledge and information from soil data.

A. Mucherino and G. Ruß. Provided recent updates on data mining and agriculture. This analysis focused on the problem of identifying problematic wine fermentations and predicting yield production.

Jay Gholap,AnuragIngole et al has done the research on analysis of soil dataset using data mining techniques.

VII. Data Analysis and Hypothesis Testing using WEKA

Production of Tur as kharip crop in Akola District: Production of Tur as crop in Akola District is mainly depends on Temperature followed by Rainfall and Area. Figure 2, depicts the regression analysis a data mining technique using WEKA tool

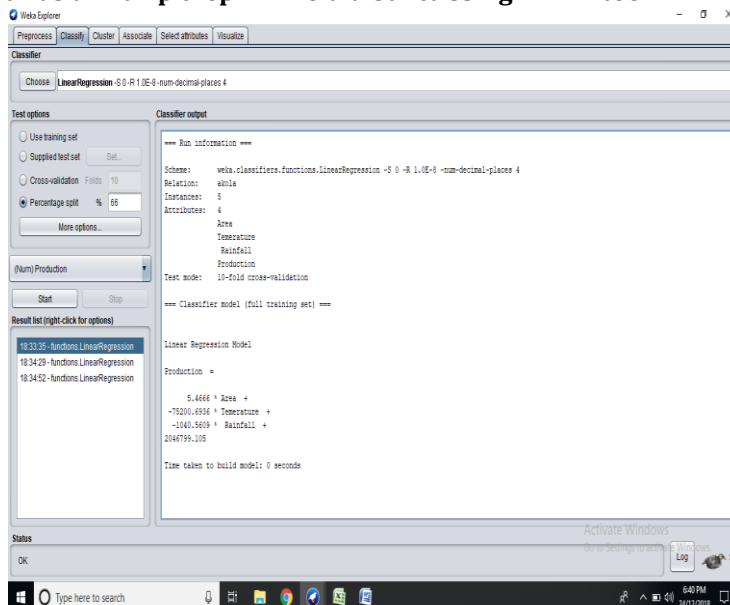**1.   Production of Tur as a kharip crop in Akola district Using WEKA tool**

Figure.2 Linear regression using WEKA tool.

i.   Linear Regression Model using WEKA

Production =   5.4666 * Area + -75200.6936 * Temerature + -1040.5609 * Rainfall +2046799.105
Time taken to build model: 0 seconds

1. From the above regression equation, researcher concludes that, Production of tur in Akola district mainly depend on temperature followed by Area and rainfall.
2. The   average Production of tur in Akola district is 76729 tones and average kharip season temperature is 28 degree Celsius and kharip season rainfall is 126.92 mm.
3. If temperature is increase by 1 degree Celsius then Production of tur in Akola district will be decrease by 75200.6936 tones while other two factors are constant.
4. If area is increase by 1 hector then Production of tur in Akola district will be increase by 5.4666 tons while other two factors are constant.
5. If rainfall is increase by 1mm then Production of tur in Akola district will be decrease by 1040.5609 tons while other two factors are constant. Hence researcher suggests to Metrological Department of Maharashtra that they should predict the rainfall in Akola district in advance and accordingly inform the farmers whether they should cultivate tur or not. If predicted rainfall would be 2-5 mm more than 126.92 mm, then farmers should not cultivate tur in that season.

Conclusion

This research study is based on the regression analysis a data mining techniques using WEKA. This research considers the proper selection of linear regression analysis a data mining technique to improve the Production of tur as a kharip crop in Akola distrct of Maharashtra. In this study, researcher concludes that Production of tur in Akola district will be depends on temperature followed by area and rainfall, if temperature in Akola Will be increase than average temperature, then Production of tur will be decrease. And if rainfall will increase than average rainfall of the season then Production of tur will be decrease. But if area will increase then Production of tur will be increase. Hence researcher suggests to Metrological Department of Maharashtra that they should predict the rainfall and temperature in Akola district in advance and accordingly inform the farmers whether they should cultivate tur or not. If predicted rainfall and temperature would be more than average, then farmers should not cultivate tur in that season.

References

Journal Articles

1. Mr. Abhishek B. Mankar, Mr. Mayur S. Burange   "Data Mining - An Evolutionary View of Agriculture" International Journal of Application or Innovation in Engineering & Management (IJAIEM)Volume 3, Issue 3, March 2014 ISSN 2319 – 4847.
2. Raorane A.A., Kulkarni R.V. "Review- Role of Data Mining in Agriculture" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (2), 2013, 270 – 272, ISSN-0975-9646.
3. D Ramesh , B Vishnu Vardhan " Data Mining techniques to estimate crop yield analysis with existing data" ISSN [Online]:-2278-1021, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.
4. S.S.Baskar, L.Arockiam, S.Charles "Applying Data Mining Techniques on Soil Fertility Prediction" International Journal of Computer Applications Technology and Research Volume 2– Issue 6, 660 - 662, 2013.
5. Shweta Taneja, Rashmi Arora,Savneet Kaur "Mining of Soil Data Using Unsupervised Learning Technique" International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 7 No.11 (2012).
6. J. Rajendra Prasad, P. Ravi Prakash, S. Sai Kumar, M. Sundara Babu K. Swarupa Rani "Identification of Agricultural Production Areas in Andhra Pradesh",International Journal of Engineering and Innovative Technology (IJEIT)Volume 2, Issue 2, August 2012.
7. 1S.Veenadhari, 2Dr. Bharat Misra, 3Dr. CD Singh" Data mining Techniques for Predicting Crop Productivity – A review article" IJCST Vol. 2, Issue 1, March 2011 ISSN : 2229 - 4333 (Print)|ISSN:0976-8491 (Online)
8. P. Bhargavi , Dr. S. Jyothi."Data Mining Classification Techniques on Soil Fertility Prediction", International Journal of Engineering Trends and Technology, July to Aug issue 2011, ISSN: 2231-5881.
9. Ramesh Vamanan ,K.Ramar. "Classification of Agriculture Land Soil and Data Mining Approach" International Journal on Computer Science and Engineering (IJCSE) ISSN: 0975-3397 Vol. 3 No. 1 Jan 2011.
10. Dr. D. Ashok Kumar, N. Kannathasan. "A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining" IJCSI International Journal of Computer Science Issues   published by Vol. 8, Issue 3, No. 1, May 2011, ISSN[online]:- 1694-0814.
11. A. Mucherino,and G. Ruß "Recent Developments in Data Mining and Agriculture".
12. Jay Gholap, AnuragIngole et al "Soil Data Analysis using classification Techniques and Soil Attribute Prediction".

Websites

1. www.mahaagree.gov.in
2. https://www.tutorialride.com/data-mining/regression-in-data-mining.htm
3. http://www.imd.gov.in/pages/services_agrimet.php
4. www.analyticsvidhya.com