A Detail study of production of Tur crop in Ahamednagar District using Regression Analysis a Data Mining technique

Asst.Prof. Priyanka Thakare[1] , Prof.Dr. R.M.Patil [2]
[1]Phd Regd. Student in Savitribai Phule Pune University.
[2] DYPIMR, Pune
[1]Email-priyanka.thakare3@gmail.com
[2]Email-patil.rm10@gmail.com

Abstract

Today, India ranks second worldwide in Agriculture. Maharashtra is one of the biggest, wealthiest and most developed states in India by current economic status. Economy of Maharashtra is basically depending on agriculture sector which have tremendous 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 paper focuses on a detail study of tur as a kharip crop in Ahamednagar district using regression analysis. For this the agro climatic data of Ahemadnagar district collected from the websites of Indian Meteorological department and department of agriculture Maharashtra state. This research mainly focuses on mining the large amount of tur crop data and agro climatic data using regression analysis a data mining techniques and result of the analysis is present in this article.

Keywords: Data mining in agriculture, regression Analysis.

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. Economy 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 Ahemadnagar 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 yield 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 regression a data mining technique. In the present research, this regression analysis is used to analyze the agro climatic data of Ahemadnagar districts of Maharashtra which will predict the production of tur crop.

II. Regression Analysis

Regression analysis is a statistical approach to forecasting change in a dependent variable on the basis of change in one or more independent variables. Regression involves predictor or independent variable(the values which are known) and response or dependent variable (values to predict).

The two basic types of regression are:

1. Linear regression

- It is simplest form of regression. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observe the data.
- Linear regression attempts to find the mathematical relationship between variables.
- If outcome is straight line then it is considered as linear model and if it is curved line, then it is a non linear model.
- The relationship between dependent variable is given by straight line and it has only one independent variable.
  $Y = \alpha + B X$
- Model 'Y', is a linear function of 'X'.
- The value of 'Y' increases or decreases in linear manner according to which the value of 'X' also changes.

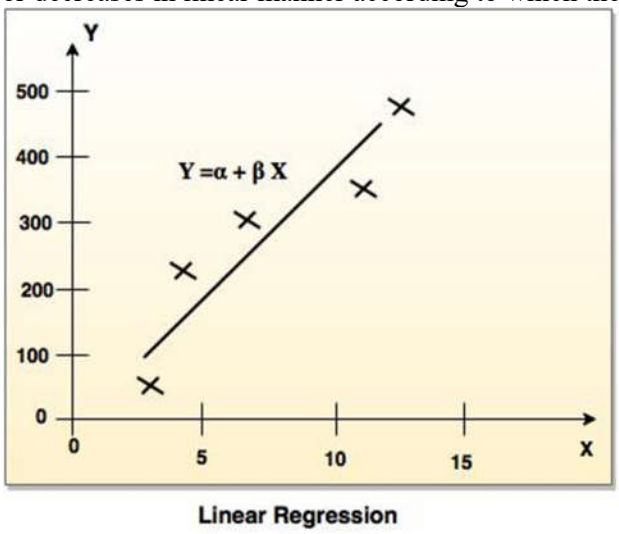

Figure.1 Linear Regression Graph

2. Multiple regression model

- Multiple linear regressions is an extension of linear regression analysis.
- It uses two or more independent variables to predict an outcome and a single continuous dependent variable.
  $Y = a0 + a1\ X1 + a2\ X2 + \ldots\ldots\ldots + ak\ Xk + e$
  where,
  'Y' is the response variable.
  X1 + X2 + Xk are the independent predictors.
  'e' is random error.
  a0, a1, a2, ak are the regression coefficients.

III. Objectives of the Study

1. To study the tur as a kharip crops in Ahemadnagar district of Maharashtra.

2. To study the agro-climatic condition in Ahemadnagar 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.

IV. Research Methodology

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

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 with following parameters namely: year, State-Maharashtra, District Ahemadnagar, crop tur, season (kharif), area (in hectares), production (in tones), average temperature (°C), average rainfall (mm).The data has been collected from Ahemadnagar 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

Researcher Shraddha Soni briefly reviewed data mining techniques with its applications and algorithms proposed. The research has done by her explained the proper selection of data mining technique and domain knowledge consideration to make effective utilization of data mining. This work will provide a pathway for beginners in this area. [1]

Tejaswini Hilage1 & R. V. Kulkarni explained how Data mining is used for mining data from databases and finding out meaningful patterns from the database. In this paper authors has reviewed the literature of data mining techniques such as Association Rules, Rule Induction Technique, Apriori Algorithm, Decision tree and Neural network. This research study focuses on how data mining techniques are used for different application areas for finding out meaningful pattern from the database.[2]

The research done by Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat Amit Khaparde described how Data mining involves extracting useful information or interesting patterns from huge historical data.In this paper researches discusses some issues in Data Mining and activities used for Data mining task.[3]

S. P. Deshpande and Dr. V. M. Thakare explained data mining applications in various fields also the different methods of data mining used to extract the patterns and thus the knowledge from this variety databases. [4]

Ashish Kumar Dogra and Tanuj Wala presented a detailed description of data mining techniques and algorithms.They describes how the various algorithms used for the mining of. [5]

Sang Jun Lee have done the detail study how to use the right information at the right time is crucial for making the right decision and this is happen by using data mining. Such as the problem of collecting data, which used to be a major concern for most organizations, is almost resolved. In the

millennium, organizations will be competing in generating information from data and not in collecting data. This study explained how data mining would be a critical factor for business success in the recent years. [6]

Shivani Goel explained, the data mining techniques can be widely classified into classification, regression and clustering. This research study describes the various applications of each data mining techniuqe. Also there are many tools available which provide methods to do different operations like WEKA, Shogun, Orange, Scikit-learn etc.This is useful for various users to select the tool best suitable for their application. All the tools do not support all the data mining operations. WEKA and Shogun supports all the three operations wiz. classification, regression and clustering while Scikit-learn supports regression and clustering operations. This research also determined which data mining tool support which technique [7]

K.Suguna and Dr.K.Nandhini explained the web mining concept and how it is use over the network.Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services.This study gives an overall idea about the data mining techniques which can be used on various server log files to find the most frequent patterns. This research study also explained data mining techniques can be used to find the user behavior over the internet [8]

Smita and Priti Sharma provide a survey of various data mining techniques. These techniques include association, correlation, clustering and neural network. This study provides a general idea of data mining, data techniques and uses data mining in various fields. [9]

Bharati M. Ramageri this research discussed few of the data mining techniques, algorithms and some of the organizations which have adapted data mining technology to improve their businesses and found excellent results. This research study explained how data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc., in different business domains. [10]

The research done by M. Durairaj, V. Ranjani had placed a focused on variety of data mining techniques, approaches and different tools and its impact on the healthcare sector. This paper aims to make a detailed study report of different types of data mining applications in the healthcare sector and to reduce the complexity of the study of the healthcare data transactions. for diagnosing or predicting diseases in healthcare sector could yield more promising results. [11]

Paper Aarti Sharma,Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava these researcher explained that the data mining is a powerful and a new field having various techniques It helps in finding the patterns to decide future trends in medical field.Data mining techniques such as classification, clustering, prediction, association and sequential patterns etc. The commercial, educational and scientific applications are increasingly dependent on these methodologies. Decision trees are a reliable and effective decision making technique which provide high classification accuracy with a simple representation of collected KDD. It helps experts to validate and classify the results and outcomes of tests and analyze various new symptoms of diseases based on data. Thus this research study concluded that, data mining can help to play an important role in the field of medicine or health care and disease prediction. [12]

Jayanthi Ranjan observed that Data mining techniques are seldom used in a pharmaceutical environment. This research described that these techniques can be easily and successfully used. The paper presented on how Data mining discovers and extracts useful patterns from this large data to find observable patterns. This study demonstrates the ability of Data Mining in improving the quality of decision making process in pharmaceutical industry. [13]

The research done by Lakshmi Prabha and Dr.A.R.Mohamed Shanavas determined education institutions are beginning to use data mining techniques for improving the services they provide and for increasing student grades and retention. This research study presents broad areas of applications in which educational data mining can be applied to e-learning. The application areas discussed in this paper are User modeling, User grouping or profiling, Domain modeling and Trend analysis. This research study discussed about the applications of EDM that made predictions and recommended actions based on increased visibility into student actions. [14]

VII. Data Analysis and Hypothesis Testing

Production of Tur as kharip crop in Ahmednagar District: Production of Tur as crop in Ahmednagar District is mainly depends on Temperature followed by Rainfall and Area.

1. **Production of Tur as a kharip crop in Amamednagar district**

Table1.Regression Analysis Coefficients

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	-127601.045	8951.866		-14.254	.045
	Area covered for tur	1.579	.031	1.003	50.687	.013
	Rainfall for Tur as Crops in Kharif season	-.746	5.740	-.003	-.130	.918
	Temperature for Tur as Crops in Kharif season	4425.477	331.951	.227	13.332	.048
a. Dependent Variable: Production of Tur as Crops in Kharif season						

Y=a+b1x1+b2x2+b3x3

x1=Area

x2=Rainfall

x3=Temperature

Y =Production

hence, Y = -127601.045 + 1.579x1 - o.746x2 + 4425.477x3

X3=4425.477

X1=1.59

X2= -0.746

1. From the above regression equation, researcher concludes that, Production of tur in Ahmednagar district mainly depend on temperature followed by Area and rainfall.

2. The average production of tur in Ahmednagar district is 7488 tones and average temperature is 26.2 degree Celsius and rainfall is 110.868 mm.

3. If temperature is increase by 1 degree Celsius then Production of tur in Ahmednagar district will be increase by 4425.477tones while other two factors are constant.

4. If area is increase by 1hector then Production of tur in Ahmednagar district will be increase by 1.579 tons while other two factors are constant.

5. If rainfall is increase by 1mm then Production of tur in Ahmednagar district will be decrease by 0.746 tons while other two factors are constant. Hence researcher suggests to Metrological Department of Maharashtra that they should predict the rainfall in Ahmednagar 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 110.868 mm, then farmers should not cultivate tur in that season.

Conclusion

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

References

Journal Articles

1. Shraddha Soni IIPS, DAVV ,A Literature Review on Data Mining and its Techniques, Volume: 5 | Issue: 6 | June 2015 | ISSN - 2249-555X.

2. Mrs. Tejaswini Abhijit Hilage1 & R. V. Kulkarni , Review Of Literature On Data Mining IJRRAS January 2012 Volumes/Vol10Issue1/IJRRAS.

3. Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat Amit Khaparde,A Review Paper on Various Data Mining Techniques Volume 4, Issue 4, April 2014 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

4. Mr. S. P. Deshpande1 and Dr. V. M. Thakare2,Data Mining System and Applications: A Review International Journal of Distributed and Parallel systems (IJDPS) Vol.1, No.1, September 2010.

5. Ashish Kumar Dogra1 , TanujWala2,A Review Paper on Data Mining Techniques and Algorithms,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015 1976 ISSN: 2278 – 1323

6. Sang Jun Lee ,A review of data mining techniques, University of Nebraska-Lincoln, Lincoln, Nebraska, USA Keng Siau University of Nebraska-Lincoln, Lincoln, Nebraska, USA

7. Mansi Gera, Shivani Goel ,Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 18, March 2015

8. K.Suguna ,Dr.K.Nandhini ,Literature Review On Data Mining Techniques ISSN:2229-6093,Vol 6 (4),583-585

9. Smita1, Priti Sharma2,Use of Data Mining in Various Field: A Survey Paper IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. V (May-Jun. 2014).
10. Bharati M. Ramageri ,Data Mining Techniques And Applications / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305

11. M. Durairaj, V. Ranjani, Data Mining Applications In Healthcare Sector: A Study International Journal Of Scientific & Technology Research Volume 2, Issue 10, October 2013 ISSN 2277-8616 29 IJSTR©2013

12. Aarti Sharma,Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava ,Application of Data Mining – A Survey Paper ISSN:0975-9646. Aarti Sharma et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2023-2025

13. Jayanthi Ranjan ,Applications Of Data Mining Techniques In Pharmaceutical Industry Journal of Theoretical and Applied Information Technology © 2005 - 2007 JATIT.

14. S. Lakshmi Prabha, Dr.A.R.Mohamed Shanavas,Educational Data Mining Applications Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No. 1, August 2014 23

Websites

15. www.mahaagree.gov.in
16. https://www.tutorialride.com/data-mining/regression-in-data-mining.htm
17. http://www.imd.gov.in/pages/services_agrimet.php