

Analysis of Crop Production Dataset using R Tool

A. Srividhya, K.Sumathi

Abstract: Agriculture is the most important factor of the Indian economy, but this vital industry which currently needs more support than any other. Agriculture is the main sector which is responsible for feeding every individual, but the people involved in it are totally neglected and last to be taken care of. With newer problems cropping up every day in the most inevitable indigenous sectors, it is high time we resort to emerging technologies for solutions. Effective Data Analysis is used to predict the production values and market rates. In this paper Agriculture data of virudhunagar district is analyzed using R Tool and Microsoft Azure in different aspects such as percentage of crop produced in various season, Production details of various crops, and crop production in different years from 1997 to 2013. The dataset is also analyzed to list out the low production crops, average production crops and high production crops using R Tool.

Keywords: Agri-Dataset Analysis, R Tool for Data Analysis

I. INTRODUCTION

In the 21st century modern farming techniques should be adopted to cater to the need of the mass, Innovative techniques to be implemented so that it will produce good results. Hence the farmers and their stake holders moving towards data-driven transformations and adopting smart farming technologies and decision making systems. The soreness in agriculture is to predict the events which will show the results and rising of production cost is the real problem for the farmers. Increasing the acre of land is not a viable solution at all to grow more crops and yield high production. Technology is an important aspect in making the better use of available space, minimizing the production cost and maximizes the productivity.

Nowadays, Technology is used to collect data from multiple hydrogenous Dataset and IoT devices mounted on the field, crop and to ground. The status of the crop is analyzed by capturing real time data from the field sensors. These data can be used for predictive analysis so that the necessary advisory services can be sent to the farmers frequently through internet, mobile devices etc., and sending necessary insights to the farmers help to improve the agriculture activities.

Agribusiness needs to improve their productivity to compensate with the growth of population at a rapid speed. Productivity analysis is used to increase the productivity over the subsequent years.

Nowadays BigData Analytics, IoT and smart agriculture can help farmers to enhance their productivity, minimize production cost and utilize the natural resources efficiently. Integrating IoT with data analytics in agriculture can help to acquire accurate predictions for crop and market conditions and also increase their yields and profits.

Revised Manuscript Received on December 16, 2019.

A. Srividhya, Department of CS & IT, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar, Tamilnadu, India,

Dr.K.Sumathi, Assistant Professor, Department of CS & IT, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar, Tamilnadu, India.,

Most of the Data analysis is carried out using R tool for statistical computing and visualization [1], R is an open-source dialect of the S statistical computing language. It is free, runs on most computing platforms, and contains contributions from top computational statisticians.

Various data analytics tools are used by the researchers to analyze the dataset. R Tool provides a variety of ways to analyze the data. R Language provides many add-on packages which have been written by other statisticians. R provides over 5,000 contributed packages which are available for free download from CRAN[2]. RStudio is an IDE in which we can run R programs.

II. RELATED WORKS

Ch. Chandra Sekhar et.al [3] applied data clustering techniques to examine disseminated dataset and proposed a Big Data Analytics Framework on Crop Sales Data and explained the crop sale data analysis steps. Reliable crop dataset collected from Open Government Data (OGD) Platform India published by National Data Sharing and Accessibility Policy. By using R Tool, data is analyzed and output that crop growing sample and diversification, excessive yield production, agro-climatic region wise performance and crop production. Narendra Gupta et. al [4] discussed about the a) importance of data analysis in agriculture b) how to store the valuable data in computerized form. They also explained that a) most of the organizations keeping their data in MS Excel, MS Word, MS Access or MS SQL and RDBMS b) a well-organized data maintaining structure have been developed by different countries

Jharna Majumdar et al. [5] implemented various data mining techniques to assess the best performance yielding method. Data mining techniques used are PAM, CLARA and DBSCAN to obtain the optimal climate requirement of wheat like optimal range of best temperature, worst temperature and rain fall to achieve higher production of wheat crop. Changes in Cropping Pattern and Crop Output and Growth Performance of Major Crops At National Level is discussed by Elumalai Kannan and Sujata Sundaram[6]

III. PROPOSED FRAMEWORK

Dataset is downloaded from tamilnadu agriculture dataset[1]. This agriculture dataset includes 13,547 records which describes the crop production details of 31 districts of tamilnadu from 1997 to 2013. This data set contains Crop Data which is collected from different districts of Tamilnadu. The dataset includes State_Name, District_Name, Crop_Year, Season, Crop, Area, Production details. In this paper, crop production data of virudhunagar district is analyzed using various machine learning tools.

Dataset is analyzed in various aspects using R tool. a) crop produced in various season b) Production details of various crops c) crop produced in different years from 1997 to 2013 . The dataset is analyzed for finding the crops that are produced frequently and

Analysis of Crop Production Dataset using R Tool

rarely using R Tool. R tool provides lot of grouping and aggregation methods for data analysis. It also provides numerous methods to visualize the output data.

The data set includes 13547 records, 493 records are extracted for analysis. Crop production data of various seasons was analyzed using R Tool. The data set includes 3 seasons namely Whole Year, Kharif, and Rabi. The productivity under these categories is analyzed. The analysis of crop produced in various season is displayed in Figure 1.

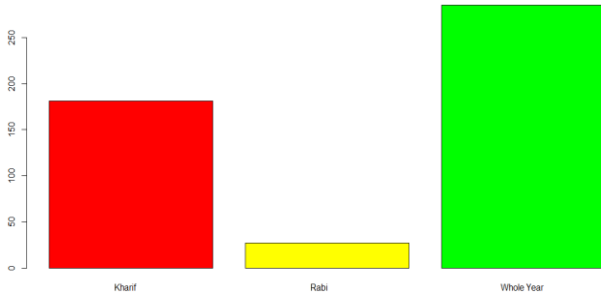


Figure 1. Analysis - crop produced in various season

The production details of various crop is analyzed. Totally the 65 crops are produced around virudhunagar district. Frequently sowing crops are a) Ragi, Groundnut, Cotton(lint) are produced in 19 farms, b) Urad, Moong(Green Gram), M maize, Sunflower crops are produced in 18 farms of different location in virudhunagar district. Rarely sowing crops are 1) Dry ginger 2) Potato 3) Other Cereals & Millets 4) Sannhamp 5) Guar seed 6) Grapes produced in only one farm.

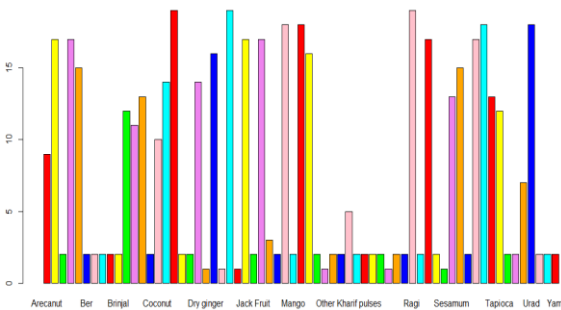


Figure 2. Analysis - Production details of various crops

The Crop production details of various years are also analyzed using R Tool. Dataset includes production details from 1997 to 2013. In the year of 2001, only 12 farmers produced 12 different crops in their farm. In 2003, 56 farmers produced 54 different crops in various seasons.

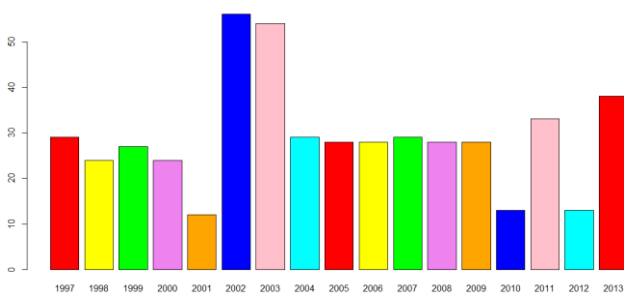


Figure 3. Analysis - crop produced from 1997-2013

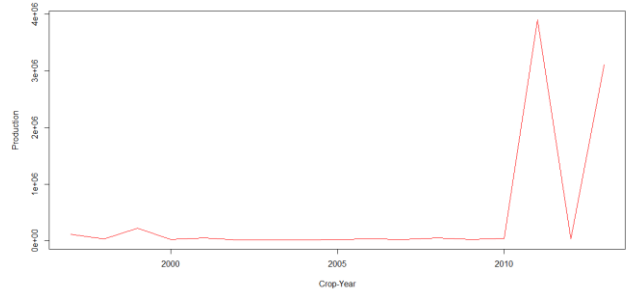
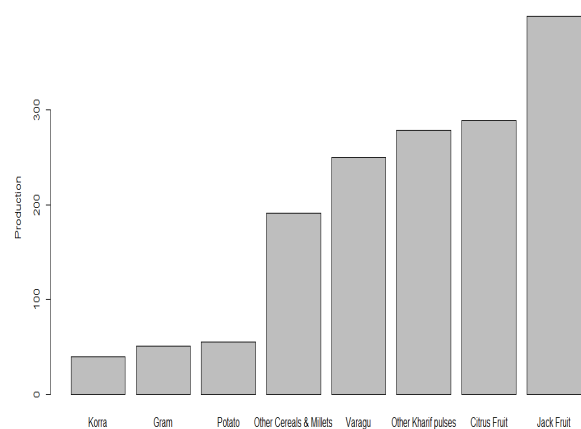


Figure 4. Analysis - Crop Year and Production

The production details of various crops are also analyzed using R Tool. Dataset includes production details 63 crops which are produced by 493 farmers. The production details of various crops for 1997 to 2013 are displayed in table 1 and figure 6. Crop wise production chart is displayed in Figure 7.

Table1 . Analysis - Crop wise Production Details

Cropname	Production
Arecanut	1.555556e+00
Arhar/Tur	4.009412e+02
Ash Gourd	0.000000e+00
Bajra	9.725118e+03
Banana	3.15E+04
Ber	0.000000e+00
Bhindi	1.030000e+03
Bitter Gourd	0.000000e+00
Bottle Gourd	0.000000e+00
Brinjal	1.993000e+03
Cardamom	2.600000e+01
Cashewnut	2.345455e+01
Castor seed	1.023077e+01
Citrus Fruit	2.890000e+02
Coconut	2.464034e+07
.	.
.	.
.	.
Tomato	2.670000e+03
Total foodgrain	8.400800e+04
Turmeric	3.371429e+01
Urad	2.668833e+03
Varagu	2.500000e+02
Water Melon	0.000000e+00
Yam	0.000000e+00



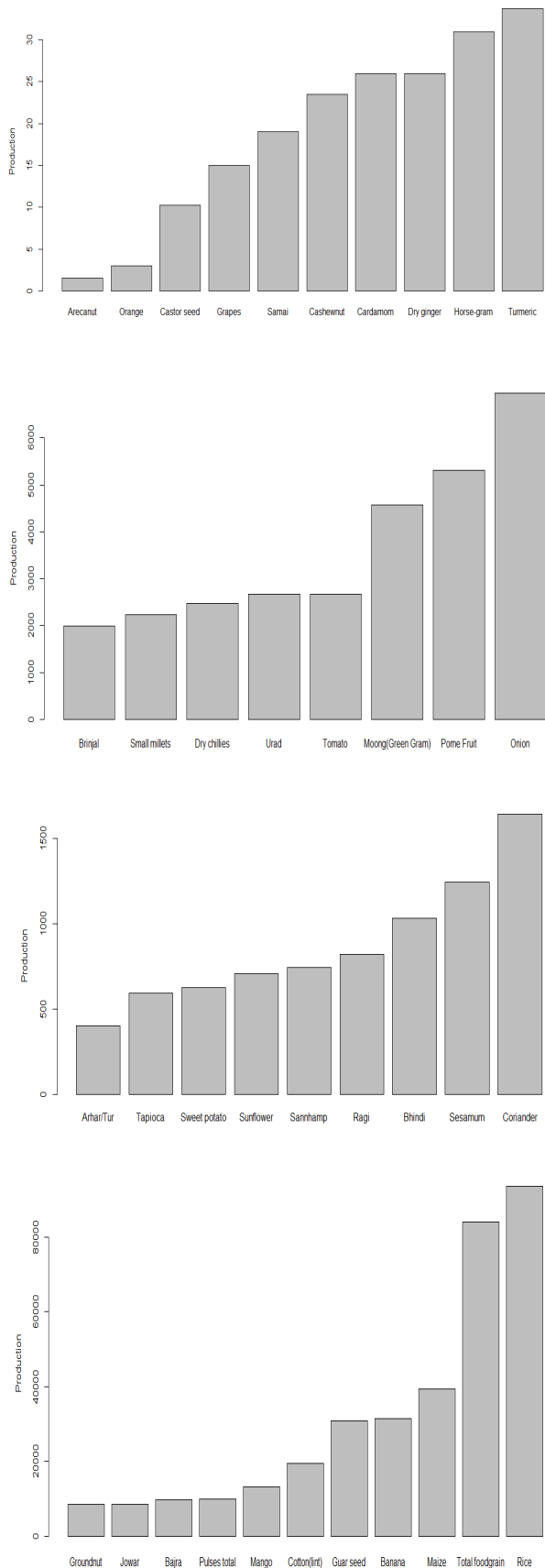


Figure 6. Crop wise production - Analysis

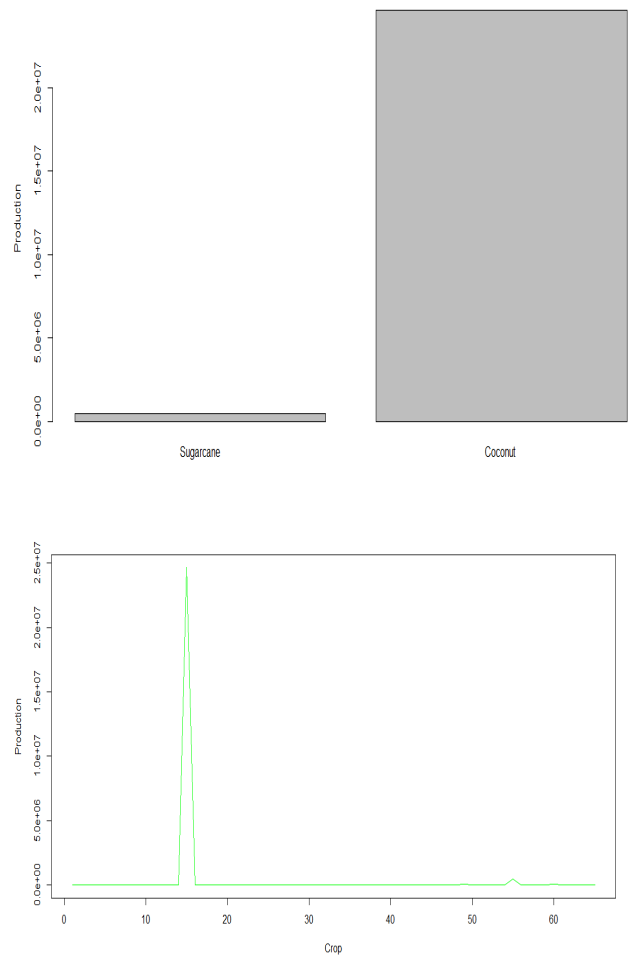


Figure 7. Crop wise production Chart

The dataset is analyzed and crop names are listed out under high production, low production and average Production in table 2, table 3 and table 4 respectively.

Table 2 - High Production Crops

Crop	Production
Mango	1.322800e+04
Maize	3.938783e+04
Coconut	2.464034e+07
Cotton(lint)	1.941379e+04
Total foodgrain	8.400800e+04
Guar seed	3.087100e+04
Rice	9.346894e+04
Banana	3.145773e+04
Sugarcane	4.973585e+05

Table 3 - Low Production Crops

Crop	Production
Grapes	1.500000e+01
Dry ginger	2.600000e+01
Gram	5.137500e+01

Analysis of Crop Production Dataset using R Tool

Orange	1.555556e+00
Samai	1.900000e+01
Horse-gram	3.094118e+01
Potato Other Cereals & Mill	5.500000e+01
Castor seed	3.000000e+00
Cashewnut	2.345455e+01
Turmeric	3.371429e+01
Cardamom	2.600000e+01
Korra	3.966667e+01
Varagu	1.023077e+01

Pulses total	9.92E+03
--------------	----------

The productivity is reduced gradually from 2003 due to a) poor techniques and production b) Rural atmosphere is not good c) people are having very less area of land from 2 to 5 hectares c) non availability of advisory services d) lack of technology usage.

IV. CONCLUSION

In a nut-shell Effective business decisions are mainly focused on accurate agriculture data and reliable analysis. The Farmers and other stakeholders have started using technology to get benefits such as cost savings, reduction in fertilizer, yield optimization, etc., In this paper, Agriculture data of virudunagar district is analyzed using Microsoft Azure and R Tool in different aspects such as percentage of crop produced in various season, Production details of various crops, crop production in different years from 1997 to 2013. The dataset is also analyzed to list out the low production crops, average production crops and high production crops using R Tool.

REFERENCES

- [1] R Ihaka and R Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996. 1
- [2] <https://rstats4ag.org/intro.html>
- [3] Ch. Chandra Sekhar¹, J. UdayKumar¹, B. Kishor Kumar¹ and Ch. Sekhar², Effective use of Big Data Analytics in Crop planning to increase Agriculture Production in India, *International Journal of Advanced Science and Technology* Vol.113 (2018), pp.31-40
- [4] Gupta N.K. Isaac R.K and R.K. Singh, Maintenance and Analysis of Agricultural Data: A Challenge, *INTERNATIONAL JOURNAL OF BIOASSAYS* ISSN: 2278-778X. pg. 4842-4848
- [5] Jharna Majumdar*, Sneha Naraseyappa and Shilpa Ankalak ,Analysis of agriculture data using data mining techniques: application of big data, *journal of BigData*,2017, Pg. no- 1-15
- [6] <http://www.environmentportal.in/files/file/Analysis%20of%20Trends%20in%20India%E2%80%99s%20Agricultural%20Growth.pdf>
- [7]. <https://www.kaggle.com/vijayaghanapathy/tamil-nadu-agriculture-data-set>
- [8]. <https://arxiv.org/ftp/arxiv/papers/1310/1310.4647.pdf>
- [9]. <http://www.ijesrt.com/issues%20pdf%20file/Archive-2016/August-2016/26.pdf>
- [10] <http://ijsetr.org/wp-content/uploads/2017/04/IJSETR-VOL-6-ISSUE-4-504-507.pdf>

Table 4 - Average Production Crops

Crop	Production
Other Kharif pulses	2.78E+02
Tapioca	5.92E+02
Ragi	8.19E+02
Brinjal	1.99E+03
Tomato	2.67E+03
Groundnut	8.49E+03
Citrus Fruit	2.89E+02
Sweet potato	6.24E+02
Bhindi	1.03E+03
Small millets	2.23E+03
Moong(Green Gram)	4.58E+03
Jowar	8.61E+03
Jack Fruit	3.99E+02
Sunflower	7.05E+02
Sesamum	1.24E+03
Dry chillies	2.48E+03
Pome Fruit	5.31E+03
Bajra	9.73E+03
Arhar/Tur	4.01E+02
Sannhamp	7.44E+02
Coriander	1.64E+03
Urad	2.67E+03
Onion	6.95E+03