Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.8 No.S3, 2019, pp. 154-158

Classification of Pests for Rice Crop Using Big Data Analytics

R. P. L. Durgabai[1], P. Bhargavi[2] and S. Jyothi[3]
[1,2&3]Department of Computer Science, Sri Padmavati Mahila Visva Vidyalayam, Andhra Pradesh, India
E-Mail: poonamramchandra@gmail.com, pbhargavi18@yahoo.co.in, jyothi.spmvv@gmail.com

***Abstract* - Data, in today's world, is essential. The Big Data technology is rising to examine the data to make fast insight and strategic decisions. Big data refers to the facility to assemble and examine the vast amounts of data that is being generated by different departments working directly or indirectly involved in agriculture. Due to lack of resources the pest analysis of rice crop is in poor condition which effects the production. In Andhra Pradesh rice is cultivated in almost all the districts. The goal is to provide better solutions for finding pest attack conditions in all districts using Big Data Analytics and to make better decisions on high productivity of rice crop in Andhra Pradesh.**
***Keywords:* Rice Crop, Pest, Production, Big Data, Agriculture**

I. INTRODUCTION

The new technology that goes by the name Big Data has already made influence in other industries from IT to healthcare. And now, investors are setting up to pull the potential of Big Data for the advantage of agriculture in India. In the agriculture industry big data analytics has made such an extensive impact that it is tough to identify all its effects and harder still to expect what change it might bring[1].In India Rice is cultivated in Kharif and Rabi seasons. Rice is one of the most cultivated grain crop in India and in Asian Countries.

South India consumes more rice than any part of India. Agriculture is the main profession of about 62 per cent of the people in Andhra Pradesh. Rice is the major food crop of the state contributing about 77 per cent of the food grain production. Rice is largely grown in East Godavari district and it is known as the rice bowl of Andhra Pradesh. Rice crop has been the main source of state's economy [2]. Rice Crop is attacked by different pests in various stages in all districts; this paper is to provide better solutions for finding pest attack conditions in all districts using Big Data Analytics and to make better decisions on high productivity of rice crop in Andhra Pradesh [3].

II. BIG DATA

The Big Data analytics is absolutely a revolution in the field of Information Technology. The use of data analytics by the companies is attractive every year. The company's chief focus is on customers therefore the field is successful in Business to Consumer (B2C) applications. The emergence and growth of big data analytics is essential to examine large and various data sets and to uncover hidden patterns, customer preferences and other useful information which can help organizations to make more informed business decisions. Traditional data ware houses are not suitable for unstructured and semi-structured data types as they are based on relational databases to structured data sets. Big Data is reorganized regularly and repeatedly, as in the case of real time data on stock trading, the online activities of website visitors or the performance of mobile applications the traditional databases and data ware house may not able to hold the data and processing demands of big data[4].As a result, many organizations that collect, process and analyze big data turn to NoSQL databases as well as Hadoop and its components [5], including

TABLE I HADOOP COMPONENTS

Components	Description
YARN	A cluster management technology and one of the key features in second-generation Hadoop.
MapReduce	A software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers.
SPARK	An open-source parallel processing framework that enables users to run large-scale data analytics applications across clustered systems.
Hbase	A column-oriented key/value data store built to run on top of the Hadoop Distributed File System (HDFS).
Hive	An open-source data warehouse system for querying and analyzing large datasets stored in Hadoop files.
HDFS	Hadoop Distributed file System.
Pig	Anopen-source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs to be executed on Hadoop clusters.
Flume, Sqoop	Data ingesting services.
Zoo Keeper	Managing cluster.

A. Apache Pig: Pig supports pig Latin language, which has SQL like command, structure. Pig is classified in two parts: Pig Latin the language and the pig runtime for the execution environment. Pig is useful to create query execution routines for analyzing large, distributed data sets without having to do low-level work in Map reduce, much like the way the Apache Hive data ware house provides a SQL like interface for Hadoop that doesn't require direct Map reduce Programming.

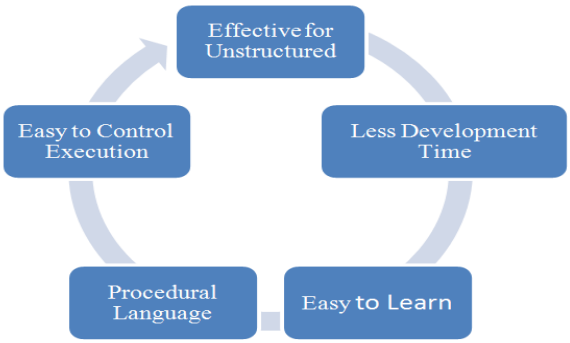

Fig. 1 Advantages of Apache Pig

Pig is having various advantages like it consumes less time while development, it is easy to learn, anyone who do not know how to write java code or map reduce programs can also learn apache pig. As apache pig is not like a declarative language and it is a Procedural language which makes user to follow the commands easily. Pig defines about data; the importance is given to the data so it can be considered as a best data flow language. It is possible to write user defined functions in apache pig, the evaluation of pig is considered lazy unless a user do not produce an output file or does not output any message. Usage of Hadoop features is a major benefit as it allows parallelization, fault tolerance with many relational database features. Pig is quite effective for large unstructured data sets; it is one of the best tools to make unstructured datasets into structured data sets[6].

III. EXPERIMENTAL SET UP AND RESULTS

In this experiment setup, edureka virtual machine Hadoop 2.7.3 infrastructure is downloaded with configuration single node having 6GB RAM and i5 CPU. Data is collected from Acharya N.G. Ranga Agriculture University and Agriculture Department of AP Government. In Andhra Pradesh during all seasons Paddy is a crop which is mainly cultivated in all districts. The present Dataset consists about all districts of Andhra Pradesh where paddy is cultivated and pest name, stage of the crop in which pest is attacked, insecticide which is given to control the pest, dosage of pesticide to the crop, after the attack of pest how much the rice is produced in each districts in the years 2015, 2016, 2017. Apache Pig is used to analyze the data and queries have been generated. Pig programs can be executed in three methods. User has to develop a pig script file containing all the commands and execute it from the command line. Grunt is an interactive shell that can be used to execute the commands line by line.

The required commands can be executed by extending the PigRunner class. The method applied is interactive shell grunt to execute commands line by line [14].Now take a query to load the rice production and pest attack data, first open the terminal and give pig -x local to store the data in a local mode and when the grunt shell is open give the following command to load the AP rice crop pest and production data.

Paddy= load'/home/edureka/desktop/pest.txt' using pig storage ('/t');
paddy_prod= load'/home/edureka/desktop/production.txt' using pig storage ('/t');

To find total production of kharif between 2016-2017
Grunt>total_kharif = foreachgroup_prodgenerate sum(rice_prod.kharif_2016_2017);

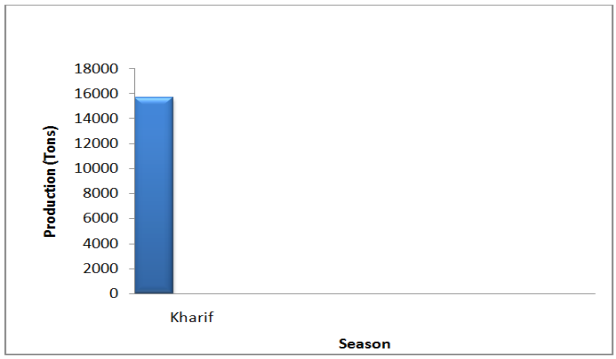

Fig. 2 Production of kharif between 2016-2017

To find total production of both kharif and rabi between 2015-2016
Grunt>total_kb = foreachgroup_prod generate sum(rice_prod.Total_2015_2016);

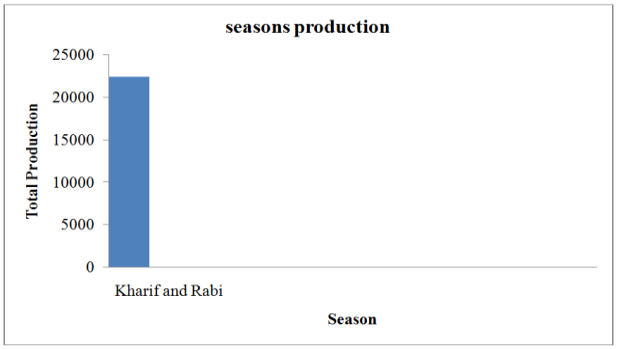

Fig. 3 Production of both kharif and rabi between 2015-2016

To find max production of rabi between 2016-2017
Grunt>max_prod = foreachgroup_prod generate MAX(rice_prod.Rabi_2016_2017);

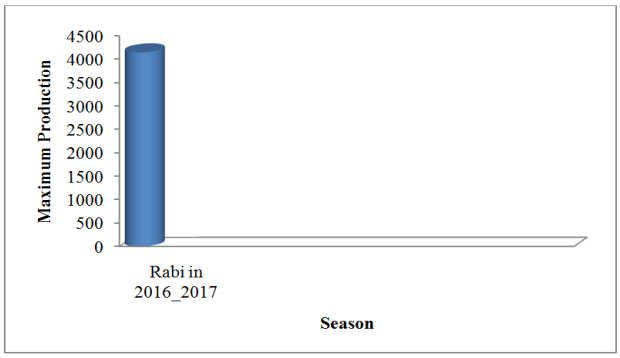

Fig. 4 Maximum production of rabi between 2016-2017

To find max of total production between 2016-2017
Grunt>total_maxprod = foreachgroup_prod generate MAX (rice_prod. Total_2016_2017);

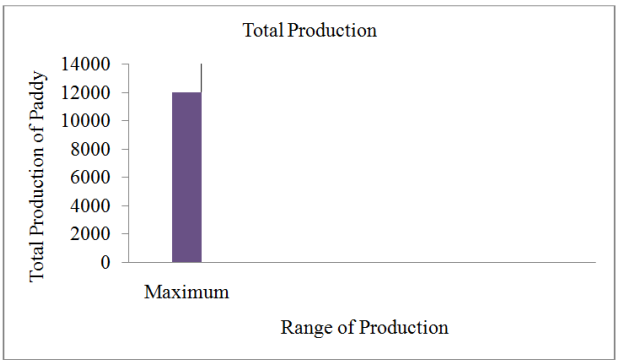

Fig. 5 Maximum of total production between 2016-2017

To find out the districts where the crops are in Tillering stage
Grunt>tillering_stage = filter P1 by Stage_Of_Crop == 'Tillering';

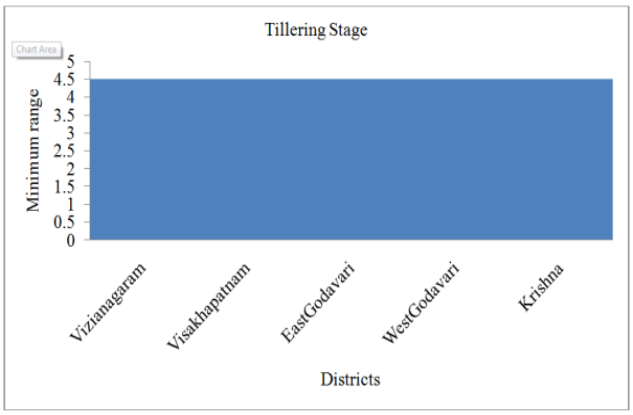

Fig. 6 Districts where the crops are in Tillering stage

To find out the districts where the crops are in Tillering stage during rainy season
Grunt> grunt> a = filter P1 by Stage_Of_Crop == 'Tillering' AND Seasonal_Condition == 'rainy_season';

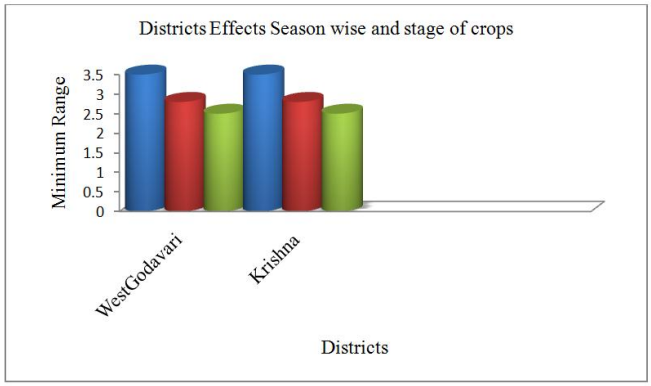

Fig. 7 Districts where the crops are in Tillering stage during rainy season

To find out the pests for which Profenophos_50_EC is being used
Grunt> c = filter P1 by Insecticide == 'Profenophos_50_EC';

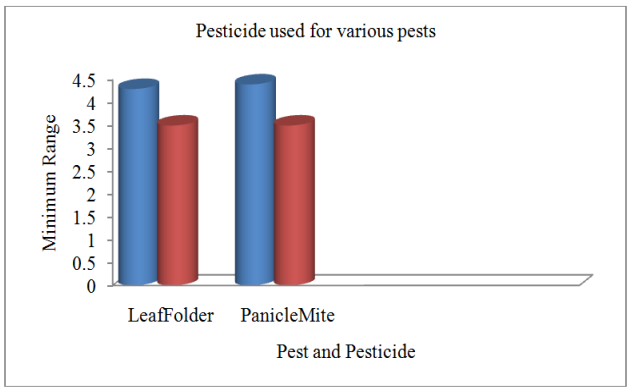

Fig. 8 Pests for which Profenophos_50_EC is being used

To find out the Districts where the seasonal condition is rainy
Grunt> b = for each a generate District, Seasonal_Condition, Stage_Of_Crop;

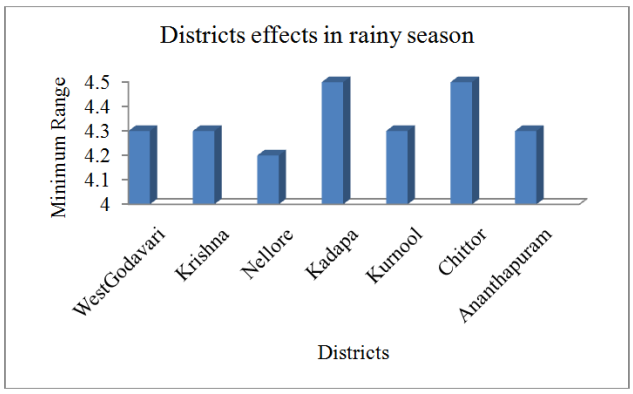

Fig. 9 Districts where the seasonal condition is rainy

To find the number of times each insecticide has been used

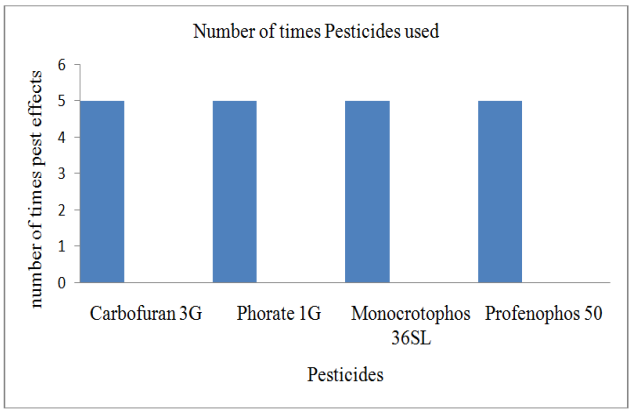

Fig. 10 Number of times each insecticide has been used

To find out the districts that is being affected by Stemborer

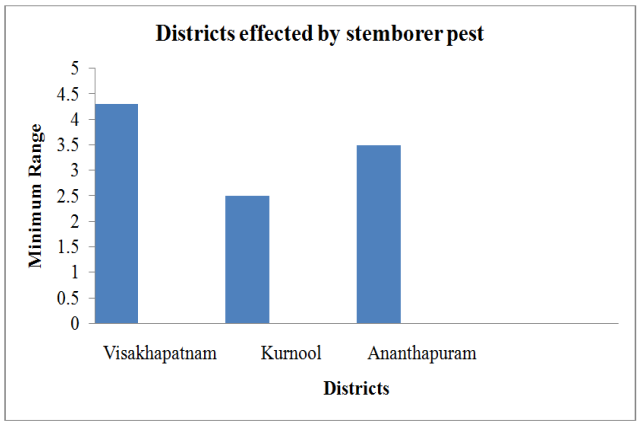

Fig. 11 Districts that are being effected by Stemborer pest

To find out the districts that is being effected by GallMidge
Grunt>district_g_1 = foreachdistrict_gallmidge generate District,Pest;

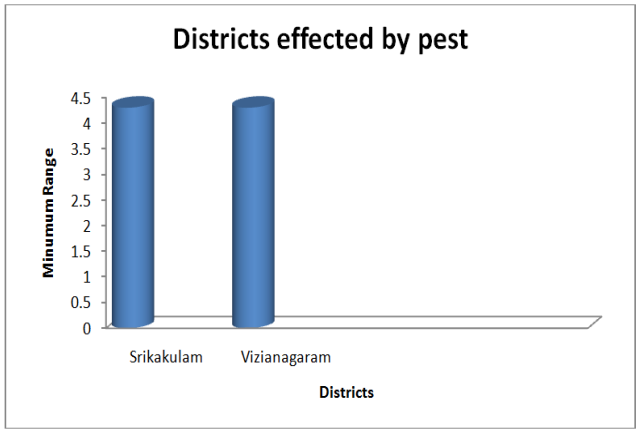

Fig. 12 Districts that are being effected by GallMidge Pest

To find out the seasonal condition,districtsin which gall midge effects the crops

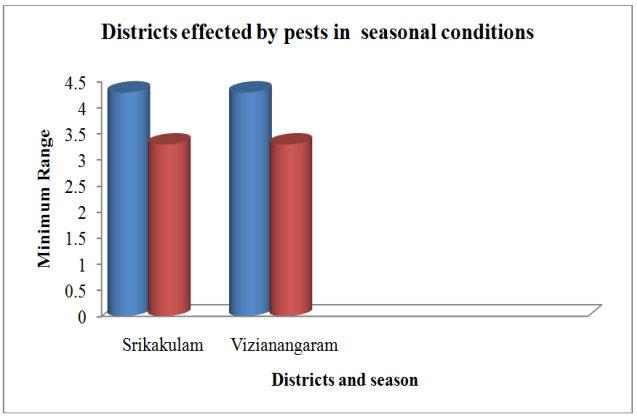

Fig. 13 Seasonal condition, districts in which gallmidge effects the crops

To find out the seasonal conditions, Districts when Stem borer are effecting the crops

Grunt>district_S_1 = foreachdistrict_stemborer generate District,Pest;

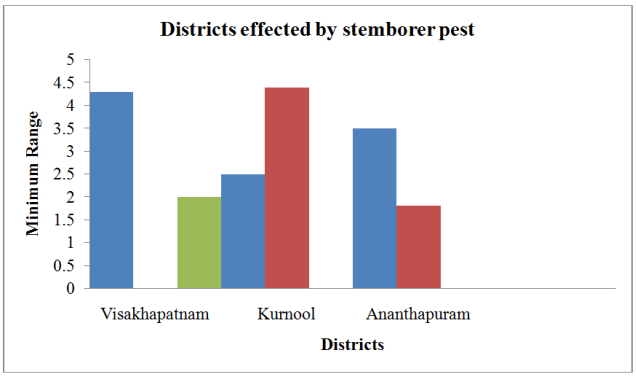

Fig. 14 Seasonal Conditions, Districts when Stem Borer are effecting the crops

IV. CONCLUSION

Agriculture has adopted a new technology so agriculture data is playing a dominant role in the world of big data. The main essence of big data is analyzing the data, which is a useful technique that helps industries to make wise decisions. To analyze the rice crop pests which is attacked by different types of pests in all districts of Andhra Pradesh, big data analytics framework has been developed that provide wide solutions using Apache Pig. Based on the evidence from original data the proposed work is helpful for researchers and scientists for recommending solutions, the work is helpful to make decisions by understanding the total rice production of each district in previous years and the pest attack in each district based on the stage of crop and which pesticides are used and how much dosage is given to control the pests.

REFERENCES

[1] Data Science in the Indian Agriculture Industry, [Online] Retrieved from https://www.analyticsvidhya.com/blog/2018/05/data-analytics-in-the-indian-agriculture-industry/ 2018

[2] N. Yogeshwara Sastry, "Agricultural Statistics at a Glance-2016-2017", *Government of Andhra Pradesh, May 2017.*

[3] Ananya Chakraborty, and Emmanuel Vijayanand Murray, "Rice Production and Productivity in Andhra Pradesh", *Research Gate, June 2011.*

[4] Mukeshkumar and Mayura Nagar," Big Data Analytics in Agriculture and distribution channel", *International Conference on computing Methodologies and Communication, Accession number:17575272, July 2017.*

[5] Hadoop and its components. [Online] Available at: https://www. tutorialspoint.com/articles/apache-hadoop-and-its-components, 2015

[6] Apache Pig advantages and disadvantages, [Online] Available at: https://data-flair.training/blogs/pig-advantages-and-disadvantages/,2018

[7] Krantibansal and Priyanka Chawla, "A study of Big Data Analysis using Apache Pig", *International Journal for IJCTA*, pp. 8665-8672.

[8] C. Swarna, and Zahid Ansari, "Apache Pig-A Data flow Frame Work based on Hadoop Map Reduce", *International Journal of Engineering Trends and Technologies*, Vol. 50, 2017.

[9] SeemaAcharya, and Subhashini Chellapan, Big Data and Analytics – *Wiley Publications, 2015.*

[10] Dr. BirendraGoswami, and Pradip Kumar Chandra, "The Evolution of Big Data as A Research and Development", *International Journal*

of Scientific Research and Engineering Studies (IJSRES), Vol. 2, No. 3, March 2015.

[11] Agarwal,Shafali, and ZebaKhanam, "Map Reduce: A Survey Paper on Recent Expansion", *International Journal of Advanced Computer Science and Applications*, Vol. 6, No. 8, pp. 209-215,2015.

[12] S. Bhosale, Harshawardhan, and Devender P.Gadekar, "A Review Paper on Big Data and Hadoop", *International Journal of Scientific and Research Publications*, 2014.

[13] Chavan, Vibhavari, and Rajesh N. Phursule, "Survey paper on Big Data", *International Journal of Computer Science Information Technology*, Vol. 5, No. 6, 2014.

[14] Samak, Taghrid, Daniel Gunter, and Valerie Hendrix, "Scalable analysis of network measurements with Hadoop and Pig", *Network Operations and Management Symposium, IEEE*, 2012.

[15] N.G. Yethiraj, and Noor Ayesha, "A study to improve crop Yield in Agriculture using IOT and Bigdata", *Adarsh Journal of Information Technology*,Vol. 6,2017.

[16] D.Laney, "3D data management: Controlling data volume, velocity and variety". *Meta Group Inc Application Delivery Strategies*, 2012.

[17] X.W.Chen, X. Lin, "Big data deep learning challenges and perspective", *IEEE Access*. Vol. 2, pp. 514–22, 2014.

[18] V.Marx, "Biology: *The big challenges of big data Nature*", Vol. 498, No. 7453, pp. 255–60,2013.

[19] Big Data in Agriculture [online] Available at:http://www.citethis-forme.com/topic-ideas/technology/'Big%20Data'-6678234.

[20] H. Zhang, X. Wei, T. Zou, Z.Li, and G. Yang, "Agriculture big data: Research status, challenges and countermeasures", *Proceedings of Computer and Computing Technologies in Agriculture*, China, 2014.

[21] A. Schumacher, L. Pireddu, M. Niemenmaa, A. Kallio, E. Korpelainen, G. Zanetti, and Heljanko K. Jan, "Simple and scalable scripting for large sequencing data sets in hadoop Bioinformatics", Vol.30, No. 1, 2014.

[22] S. Arjun, Anish Joshi, H.Pooja Das, and R. Amutha, "Big Data Analytics for Agriculture Development in India", *International Journal of Engineering Research and Technology*, 2016.

[23] Ehizogie Omo-Ojugo, "Relevance of Big Data Analytics in Agriculture: Focus on Nigeria Agricultural Sector", *International Journal of Science and Research*, 2017.

[24] K. Ravisankar, K. Sidhardha, and B. Prabadevi, "Analysis of Agricultural Data Using Big Data Analytics", *Journal of Chemical and Pharmaceutical Sciences*, Vol. 10, No. 3, 2017.

[25] S.S. De, G. Chattopadhyay, B. Bandyopadhyay, and S. Paul, "A neuro-computing approach to the forecasting of monthly maximum temperature over Kolkata, India using total ozone concentration as predictor", *Comptes Rendus Geoscience*, Vol. 343, No. 10, pp. 664-676, 2011.