Big Data Framework For Classification of Agriculture Land Category of Kanchipuram District Tamil Nadu

Naveen Raj. D[1],Dr.N. Ayyanathan[2]
[1]Dept of Computer Applications
[2]Associate Professor, Dept of Computer Applications
[1, 2] B. S. Abdur Rahman Crescent Institute of Science & Technology, Vandalur, Chennai-600048, India

Abstract- *In this project we are analyzing agriculture data by using Hadoop tool along with some Hadoop ecosystems like HDFS, MapReduce, Sqoop, hive and pig. By using these tools, we can process no limitation of data, no data lost problem, we can get high throughput, maintenance cost also very less and it is an open-source software, it is compatible on all the platforms since it is Java based. In agriculture data details, to finding of agricultural lands in Kanchipuram district. It also aimed to find out the areas of rapid change, magnitude of change and assess the past and present condition of agricultural Land. Agriculture is the science and art of cultivating plants and livestock. Agriculture was the key development in the rise of sedentary human civilization, whereby farming of domesticated species created food surpluses that enabled people to live in cities. Monitoring crop and natural vegetation conditions is highly relevant, particularly in the food insecure areas of the world.Data from remote sensing image time series at high temporal and medium to low spatial resolution can assist this monitoring as they provide key information about vegetation status in near real-time over large areas.Using Big Data technology, we can bring the classification results presented about 90% of correlation with official estimates based on both traditional and satellite image analysis methods. This can make the area detection be satisfied outcome.*

Keywords- Agriculture area, Time series, Agriculture land, Hadoop, HDFS.

I. INTRODUCTION

Agriculture is the science and art of cultivating plants and livestock. Agriculture was the key development in the rise of sedentary human civilization, whereby farming of domesticated species created food surpluses that enabled people to live in cities [1]. Monitoring crop and natural vegetation conditions is highly relevant, particularly in the food insecure areas of the world [2]. Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation[3]. Still the most agricultural fields are under developed due to the lack of deployment of ecosystem control technologies[4].Due to these problems, the crop production is not improved which affects the agriculture economy[5].So, by using the data analysis, we can get an insight of what the result will be and also using criteria we can make a prediction based on the analysis using R language[4]. As of now, manual framework is being used in the drug store[5].So, by using the data analysis, we can get an insight of what the result will be and also using criteria we can make a prediction and visualization based on the analysis using R language[6].

III. PROBLEM DEFINITION

The research work is being undertaken to address the finding of agricultural lands in Kanchipuram district from 1997 to 2018 as it reflects the basis for addressing production deficits in the food safety assessment. It also aimed to find out the areas of rapid change, magnitude of change and assess the past and present condition of agricultural Land to understand the dynamics and trend of change. It is very useful to government to find out the agricultural land and restore the agricultural land.

IV. LITERATURE REVIEW

Monitoring agricultural land cover is highly relevant for global early warning systems such as ASAP (Anomaly hot Spots of Agricultural Production), because it represents the basis for detecting production deficits in food security assessment [1].Given the significant inconsistencies among existing land cover datasets, there is a need to obtain a more accurate representation of the spatial distribution and extent of agricultural area [2].In this research, we explore a fusion approach that combines the strength of individual datasets and minimizes their limitations. Specifically, a semi-automatic method is developed, relying on multi-criteria analysis (MCA) complemented with manual fine-tuning using the best-rated datasets, to generate two hybrid and static agricultural masks – one for cropland and another for grassland [3]. The results indicated that there was a regular trend characterized in most classes and that the change in different land use/land cover classes ranged between increase and decrease areas [4].A continuous increase in agricultural, urban, fish farms and

natural vegetation areas and a continuous decrease in water bodies and sand areas were detected in the studied area [5].The accuracy of the classified map was assessed using a High-Resolution Planet scope image and ground realities have been verified and ascertained through field observations and site-specific interviews [6]. As a result of policy changes and traditional agroforestry systems, LULC in the study watershed has undergone a series of complicated changes over the past three decades [7].Six major LULC classes viz; agriculture, barren land, built-up area, dense forest, open forests and water bodies have been identified and indicate that major land use in the watershed is forestry [8].

V. EXISTING SYSTEM

Existing concept deals with providing backend by using MySQL which contains lot of drawbacks i.e., data limitation is that processing time is high when the data is huge and once data is lost, we cannot recover so thus we proposing concept by using Hadoop framework.

5.1 DISADVANTAGES OF EXISTING SYSTEM

- We can process limitation of data.
- We get results with take more time and maintenance cost is very high.
- The processing time is high when the data is huge and once data is lost, we cannot recover.

VI. PROPOSED SYSTEM

Proposed concept deals with providing database by using Hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster and we get results with less time, high throughput and maintenance cost is very less and we are using joins, partitions and bucketing techniques in Hadoop.

6.1 ADVANTAGES OF PROPOSED SYSTEM

- Hadoop is a highly scalable storage platform because it can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel.
- Its enables businesses to easily access new data sources and tap into different types of data (both structured and unstructured) to generate value from that data.
- Maintenance Cost is very less and also offering a cost-effective storage solution for businesses exploding data sets.
- No data loss problem.
- Efficient data processing.

VII. REQUIREMENTS SPECIFICATION

7.1 HARDWARE REQUIREMENTS

The minimum **Hardware Requirements** are,

OPERATING SYSTEM : Windows 7/ 8.1
PROCESSOR :Intel Corei5
SPEED : 2.8 MHz
MEMORY : 4 GB Ram
SDD :250GB
DISPLAY : 15” Color
KEYBOARD : 104Keys

7.2 SOFTWARE REQUIREMENTS

The minimum **Software Requirements** are,
FRAMEWORK :Hadoop
DATABASE : MYSQL 5.5
LANGUAGES :PigLatin, Core Java
DATAACCESSTOOL :Sqoop
OPERATING SYSTEM :Cent OS
IDE :Eclipse

VIII. MODULE DESIGN SPECIFICATION

8.1 PREPROCESSING DATA

In MySQL is a relational database management system. RDBMS uses relations or tables to store Agriculture data as a matrix of rows by columns with primary key. With MySQL language, Agriculture data in tables can be collected, stored, processed, retrieved, extracted and manipulated mostly for business purpose.

Figure1

8.2 CONNECTOR (SQOOP)

Sqoop is a command-line interface application for transferring Agriculture data between relational databases (MySQL) and Hadoop. Here in MySQL database having Agriculture data have to import it to HDFS using Sqoop.

Agriculture data can be moved into HDFS/Hive from MySQL and then it will generate the java classes. In previous cases, flow of data was from RDBMs to HDFS. Using "export" tool, we can import data from HDFS to RDBMs. Before performing export, Sqoop fetches table metadata from MySQL database. Thus, we first need to create a table with required metadata.

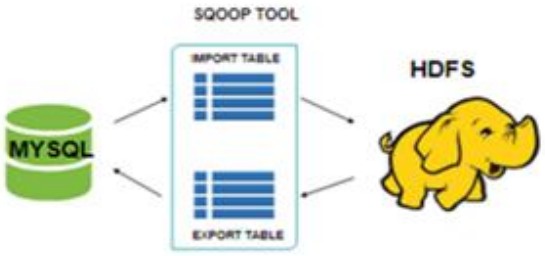

Figure2

8.33 ANALYSIS QUERY LANGUAGE (HIVE)

Hive is a data ware house system for Hadoop that runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. In Hive, Agriculture data tables and databases are created first and then data is loaded into these tables. Hive as Agriculture data warehouse designed for managing and querying only structured data that is stored in tables. Hive organizes Agriculture data tables into partitions. It is a way of dividing a table into related parts based on the values of partitioned columns. Using partition, it is easy to query a portion of the given dataset. Tables or partitions are sub-divided into buckets, to provide extra structure to the Agriculture data that may be used for more efficient querying. Bucketing works based on the value of hash function of some column of a table.

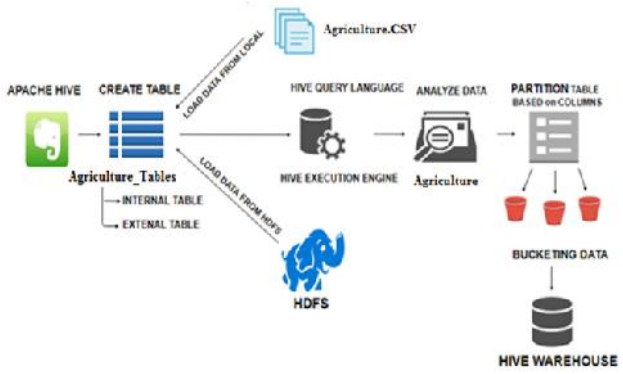

Figure3

8.4DATA VISUALIZATION

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R. We'll be using R for mostly visualizing our results.

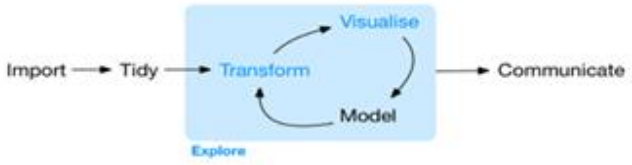

Figure4

IX. RESULTAND DISCUSSION

The result of the project **"time series analysis for agricultural land detection"** is obtained as expected. This project is implemented using Hadoop tools for the entire analysis. The tools are HDFS, Sqoop, and hive etc.…

9.1 EXPERIMENTAL SCREENSHOT OUTPUT

9.1.1 PREPROCESSING AGRICULTURE DATABASE

Figure 4

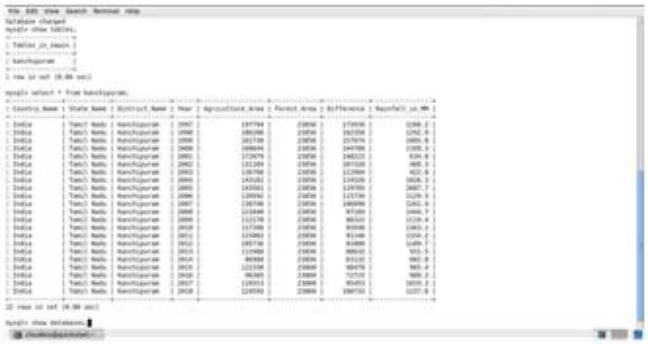

Figure 5

9.1.2 IMPORT THE TO HDFS USING SQOOP

Here in MySQL database having Kanchipuram district agriculture area data have to import it to HDFS using Sqoop.

Figure 6

Working with Eval and Querysee the table in HDFS limit5

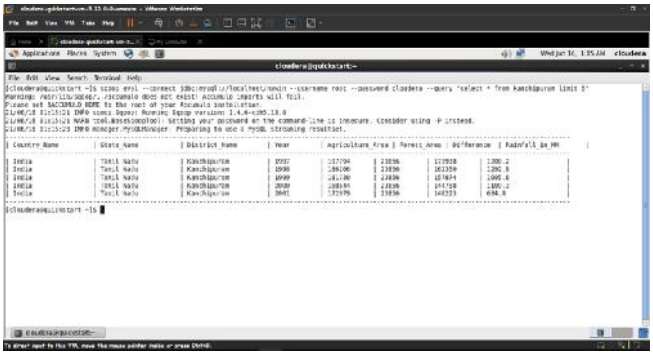

Figure 7

9.1.3 ANALYSIS QUERY LANGUAGE (HIVE)

we are getting all those data from HDFS to HIVE by use of Sqoop import command, where hive is ready to analyze. Here in HIVE, we can process only structured data to analyze by extracting only the meaningful data and neglecting unclenched data we can analyze the data in more effective manner by use of hive.

Figure 8

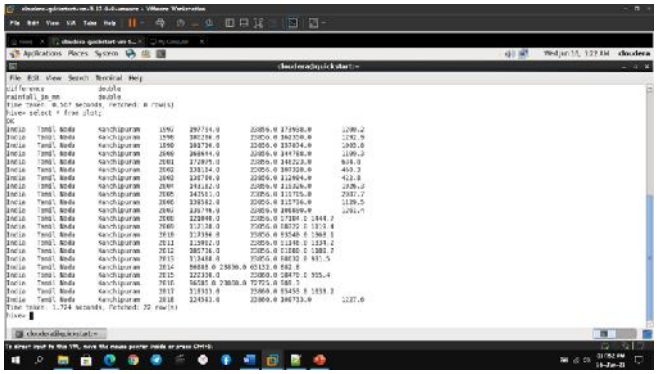

Figure 9

Agricultural area in Kanchipuram district above 150000 hectares.

Figure 10

Forest area in Kanchipuram district below 23860 hectares.

Figure 11

9.1.3 DATA VISUALIZATION IN R-STUDIO

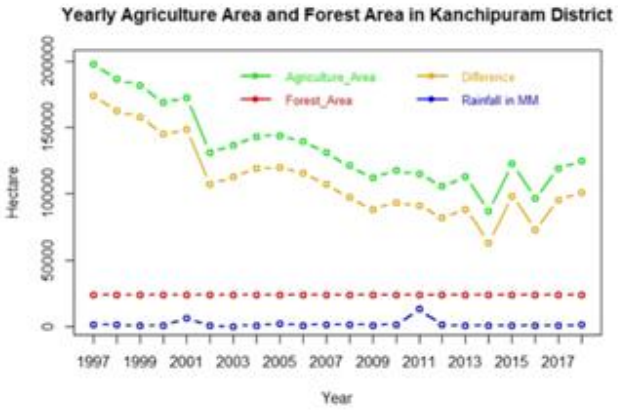

Figure 12

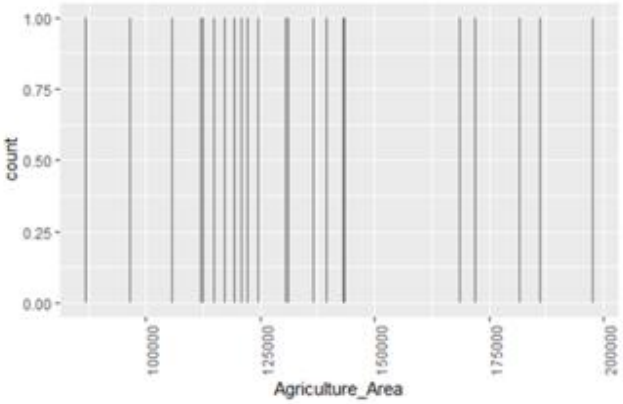

Figure 13

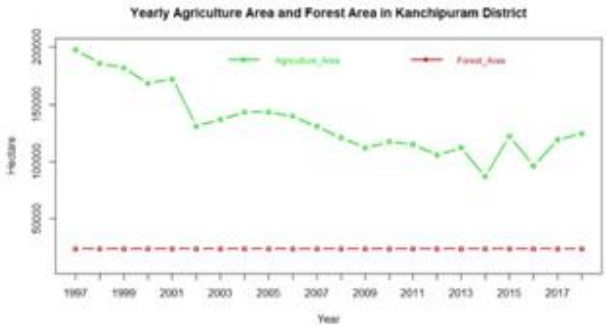

Figure 14

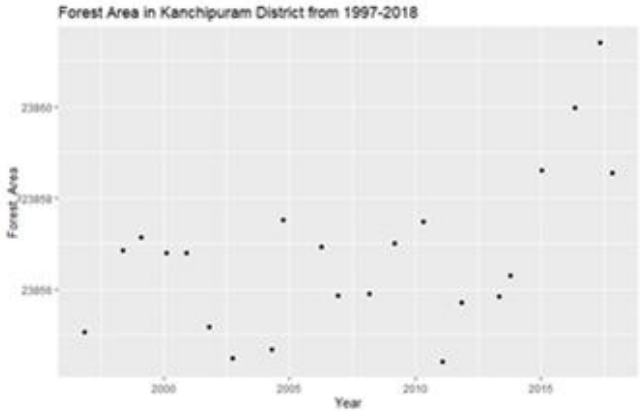

Figure 15

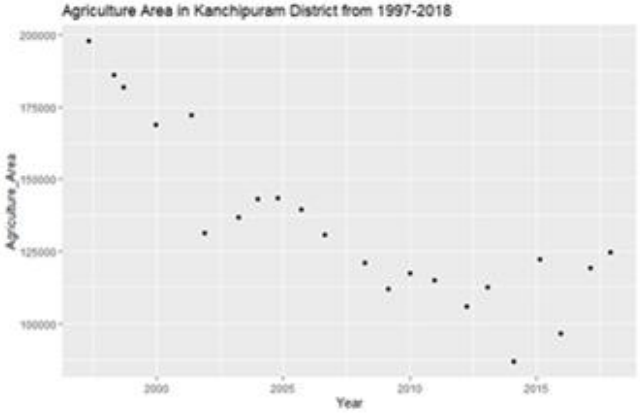

Figure 16

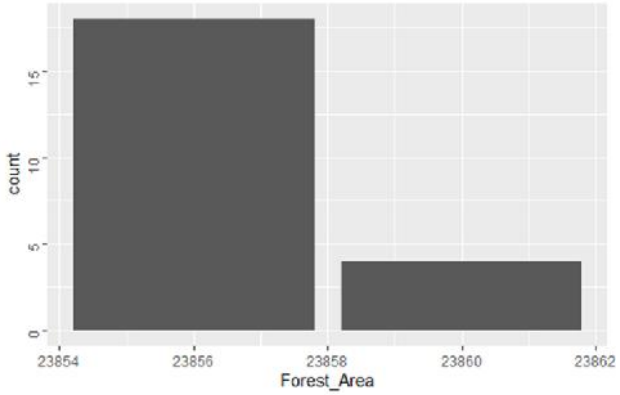

Figure 17

X. CONCLUSION

In this paper, we presented a study on Agriculture data and prediction regarding research paper about agriculture area analysis. To analysis the Agriculture data in Hadoop ecosystem and to improve the area using the soil growth, temperature, components and water level. It also aimed to find out the areas of rapid change, magnitude of change and assess the past and present condition of agricultural Land to

understand the dynamics and trend of change. It is very useful to government to find out the agricultural land and restore the agricultural land. Hadoop ecosystem is using hive, pig, map reduce tools for processing whether output will take less time to process and result will be very fast. Hence in this project, Agriculture data which is traditionally going to store in RDBMS going to less performance hence by using Hadoop tool it will be faster and efficiently processing the data.

XI. FUTURE ENHANCEMENTS

Apache Spark is an open-source processing engine built around speed, case of use, and analytics. If you have large amounts of data that requires low latency processing that a typical Map Reduce program cannot provide, Spark is the alternative. Spark provides in-memory cluster computing for lightning-fast speed and supports Java, Scala, and Python APIs for ease of development.

REFERENCES

[1] Awokuse, T.O.; Xie, R. Does agriculture really matter for economic growth in developing countries? Can. J. Agric. Econ. 2015, 63, 77–99.
[2] Gillespie, S.; Van den Bold, M. Agriculture, food systems, and nutrition: Meeting the challenge. Glob. Chall.2017, 1, 1600002.
[3] Patel, R. The long green revolution. J. Peasant Stud. 2013, 40, 1–63.
[4] Pingali, P.L. Green revolution: Impacts, limits, and the path ahead. Proc. Natl. Acad. Sci. USA 2012,109, 12302–12308.
[5] Wik, M.; Pingali, P.; Broca, S. Background Paper for the World Development Report 2008: Global Agricultural Performance: Past Trends and Future Prospects; World Bank: Washington, DC, USA, 2008.
[6] World Bank Group. Available online: https://openknowledge.worldbank.org/handle/10986/9122 (accessed on21 May 2020).
[7] Konikow, L.F. Long-term groundwater depletion in the United States. Groundwater 2015, 53, 2–9.
[8] Kleinman, P.J.; Sharpley, A.N.; McDowell, R.W.; Flaten, D.N.; Buda, A.R.; Tao, L.; Bergstrom, L.; Zhu, Q. Managingagricultural phosphorus for water quality protection: Principles for progress. Plant Soil 2011, 349, 169–182.
[9] Wen, F.; Chen, X. Evaluation of the impact of groundwater irrigation on streamflow in Nebraska. J. Hydrol.2006, 327, 603–617.
[10] Konikow, L.F.; Kendy, E. Groundwater depletion: A global problem. Hydrogeol. J. 2005, 13, 317–320.
[11] Schaefer, Mathias, and Nguyen Xuan Thinh. "Evaluation of land cover change and agricultural protection sites: a GIS and remote sensing approach for Ho chi minh city, vietnam." Heliyon 5, no. 5 (2019):
[12] Instituto Brasileiro de Geografia e Estatıstica, “Produc¸~aoagrıcola municipal,” Banco de Dados Sidra, [Online]. Available: https://bit. ly/2NN714S. 2017.
[13] Companhia Nacional do Abastecimento, “Series historicas,” Cana-de-Ac¸ucar— Area Total , [Online]. Available: https://bit.ly/ 2Y6t4IF. 2017.
[14] BFD (Bangladesh Forest Department), 2012. Ecotourism in Bangladesh: TanguarHaor. Ministry of Environment and Forest, Government of Bangladesh (accessedon 3 January 2012).
[15] Anderson, J.R., 1976. A land use and land cover classification system for use withremote sensor data (Vol. 964). US Government Printing Office.
[16] Arora, K.M., Mathur, S., 2001. Multi-source classification using artificial neuralnetwork in a rugged terrain. Geocarto Int. 16 (3), 37–44. https://doi.org/10.1080/10106040108542202.
[17] Barnsley, M.J., Møller-Jensen, L., Barr, S.L., 2001. Inferring urban land use by spatialand structural pattern recognition. Remote Sens. Urban Anal., 115–144