Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/2466
DOI URL: http://dx.doi.org/10.21474/IJAR01/2466

RESEARCH ARTICLE

SPATIAL TEMPORAL DATA MINING FOR CROP YIELD PREDICTION.

Aakunuri Manjula[1] and Dr. G. Narsimha[2].

1. Research Scholar, CSE, Affiliated To JNTUH Hyderabad, Telangana State, India.
2. Associate Professor, CSE, Affiliated To JNTUH Hyderabad, Telangana State, India.

Manuscript Info

Manuscript History

Received: 23 October 2016
Final Accepted: 21 November 2016
Published: December 2016

Key words:-
Data Mining, MapReduce programming, spatiotemporal data mining, crop yield prediction.

Abstract

Background: Crop yield prediction can help agricultural departments to have strategies for improving agriculture. Towards this end many techniques came into existence. Spatiotemporal data mining is one such solution that can be employed to achieve crop yield prediction. We extend the method employed by Cao *et al*. for leveraging the power of parallel computing of MapReduce framework. Especially we used Spatial Hadoop which is based on the new programming paradigm besides having spatial extensions. Towards this end, we proposed a framework and implemented the extended algorithm that is compatible with MapReduce programming.
Results: We collected five years data of Cotton and Maize crops of Karimnagar region of Telangana state, India. We implemented the proposed framework for crop yield prediction. Our crop yield prediction mechanism using MapReduce programming paradigm is tested using a prototype application. The proposed framework also secures data flow in the MapReduce framework. The results revealed that the proposed solution is encouraging and the error rate of prediction of Cotton and Maize is low.
Conclusions: Proposed framework was able to achieve crop yield prediction. Cotton and maize crops were used for the purpose. Results revealed that the proposed system outperformed the existing one with reduced error rate.

Introduction:-

Spatiotemporal data mining can be used to predict crop yield prediction besides other real world applications. This paper throws light into crop yield prediction using socio temporal data mining and extends it to use MapReduce framework in order to improve the prediction accuracy. Since the data for spatial data mining is very huge, big data mining can help in extracting comprehensive business intelligence. As Big Data refers to the data with characteristics of volume, velocity and variety. Processing a part of the Big Data cannot transform that into useful business intelligence. Figure 1 shows the reason why big data mining is important.

Corresponding Author:-AakunuriManjula.
Address:- Research Scholar,CSE,Affiliated To JNTUH Hyderabad, Telangana State,India India.

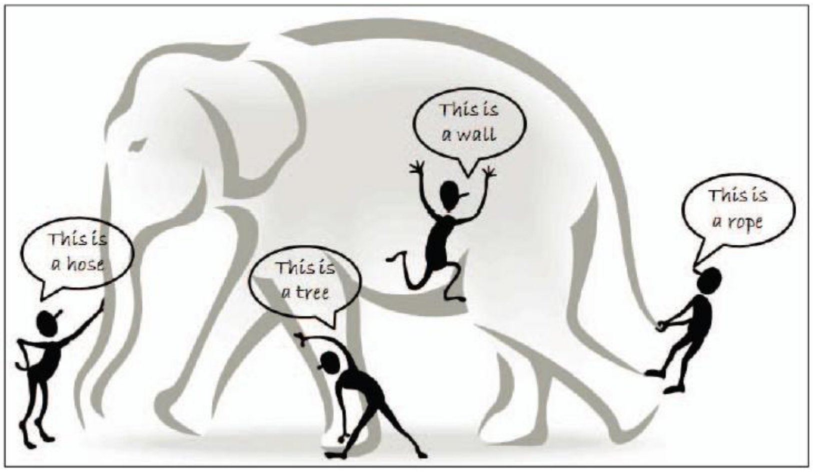

Figure 1:- The importance of big data mining.

As can be seen in Figure 1, it is evident that a part of the elephant when viewed by an individual cannot give the fact that that is elephant. Instead of revealing the truth “elephant”, user understands that part as a rope or tree etc. This analogy drives home the message that big data mining is important for obvious reasons. Big data mining only can provide actionable knowledge. Without actually mining the whole data, it provides facts which do not reflect the reality. Decisions based on such partial truths will backfire and enterprises experience huge loses. In this context it is reiterated that big data mining is important for transforming data into business intelligence.

Related works:-

Big data does mean huge amount of data that is measured generally in peta bytes (Intel, 2013). In the industry, it is a hot topic or buzz word now as it can process huge amount of data that cannot be done in the traditional environments (Ferguson, 2012). According to Chieko *et al*., 2012 Big Data mining attracted many researchers that make use of Hadoop as distributed programming framework. MapReduce is the new programming paradigm used for processing big data. This programming is done in distributed environment and it makes use of parallel processing. Over few years Hadoop became a reliable distributed programming framework. This framework is compatible with big data which is characterized by volume, velocity and variety. Hadoop is the framework that can handle such data which is static, dynamic, streaming and with various kinds. It also takes care of data correlation (Bhattacharya, 2013). Big data processing has become an essential thing for enterprises as it can transform that data into economy of business. In other words big data mining can produce reliable business intelligence and its impact on the society is so high (Levin, 2013; Schroeck et.al 2012; Economics Intelligent, 2011). Real value required by enterprises can be obtained from big data mining. To make the exercise fruitful it is important to make use of right kind of tools (Oracle White Paper, 2013).

In ITC big data processing is becoming a part as organizations need it very much (Australian Government, 2013). The phases in mining big data include data acquisition and recording, information extraction and cleaning, data representation, aggregation and integration, data modelling, analysis and query processing. Then the interpretation phase will actually produce business intelligence that supports expert decision making. Incompleteness of data, heterogeneity, human collaboration, privacy, timeliness and scale are some of the challenges pertain to big data process (Leading researchers, n.d). Financial performance of accompany can be leveraged by the effective use of big data. Exploiting data as much as possible through big data mining which can help in prioritization of activities towards achieving financial independence (SAS, 2012). Problems pertain to big data mining are experienced by companies (Jiang, 2010). NoSQL is the principle followed by the distributed programming frameworks that are involved in big data mining (Cooper, 2013). Different people think different about big data as it is different for different people. There is great scope of information processing with real time data analysis that considers huge amount of data and processing data in different angles in order to consider every aspect of business so as to have complete picture of data. As IBM said, big data is for people who get more accurate business decisions. Real time information systems can make use of big data mining so as to obtain trends or patterns that are inclusive of every minute detail for accuracy. Such knowledge can be used to make well informed decisions (Schroeck, 2012). Three Vs are associated with big data that are Volume, Velocity and Variety (Intel, 2013).

Spatial Data Mining for Agriculture:-

Jia (2009) employed spatial data mining for land grading. Kumar *et al*. (2012) proposed a method for global auto correlation pertaining to spatial data mining. Rajesh (2011) proposed spatial data mining approach for agriculture. Sharma and Mehta (2012) proposed data mining techniques to extract business intelligence from agricultural data. Julia *et al*. (2012) proposed a methodology for monitoring agricultural fields through spatiotemporal mining. Yethiraj (2012) explored data mining techniques for the field of agriculture. Bhojani (2013) explored data mining techniques for knowledge discovery in agriculture. Mucherino and Rub (2012) discussed the latest developments in data mining and the field of agriculture. Rub (2012) explored spatial data mining for precision agriculture. Hsiao *et al*. (2006) extracted coalition patterns through social data mining in order to help decision support system.

Rao *et al*. (2012) explored issues and applications in spatiotemporal data mining. Singh *et al*. (2014) studied spatio-temporal variations on earth surface air. Madhuskar and kelkar (2012) made a good review of data mining techniques. Recently Cao *et al*. (2013) proposed a method for maize yield prediction in agriculture. This method has provision for estimating various parameters and artificial neural network in order to predict crop yield accurately. In this paper we extend this method to work with MapReduce programming.

Method:-

The proposed system is meant for crop yield prediction. Two crops such as cotton and maize are considered for experiments. There are different parameters to be considered for changes in prediction of the crops. A time series model is used in order to use the data of crops and then employ a prediction model. Since prediction models have different characteristics and suitable for specific type of data, it is very important to know the choice of prediction model for improving accuracy in prediction. The yield prediction of cotton and maize belong to a spatio-temporal issue. There are many issues such as influences on the crops and land fertility. Therefore time related or temporal character of crops is considered. It does mean that both spatial and temporal domains are taken into account. The production statistics obtained from agriculture officials are used for this research.

Spatio-temporal data mining method is employed in (Ceo*et al*., 2013) in order to predict maize production. The procedure used for prediction is described here. Spatio-temporal data contains different states of an object in both time and space domains over a period of time. Environment monitoring, traffic management, weather forecasting are some of the examples for sptio-temporal data. Spatio-temporal data related to agriculture is another best example. Remote sensing data can be used to process in both space and time domain to extract patterns that can be used in making strategic decisions. Many researchers contributed towards applying spatio-temporal data mining techniques to solve real world problems. Recently Cao *et al*. (2013) proposed a spatio-temporal data mining algorithm for maize yield prediction. The outline of the algorithm is as follows.

Step 1: Defining prediction problem
Step 2: Build time series model and then build autoregressive integrated moving average (ARIMA) model from the maize yield data collected using GPS or remote sensing technology
Step 3: Use ANN to obtain hidden patterns and relationships
Step 4: Integrate spatial prediction and function and spatial temporal function and generate final spatial and temporal prediction

The problems with the approach in [1] are that the method with remote sensing or GPS data uses conventional spatial data mining technique. This technique is linear in nature and cannot perform well when huge amount of data is to be processed. It has been proved that traditional methods for spatial-temporal data mining have performance limitations. Due to the invent of new technologies like Cloud Computing, a new computing paradigm based on virtualization, and availability of parallel processing with GPUs using MapReduce programming in distributed environments, it is inevitable that the processing of huge needs to take the help of massive programming frameworks such as Hadoop (a MapReduce framework). The huge amount of data in agriculture and environment monitoring can be termed as Big Data which is characterized by Volume, Velocity and Variety. Handling such data with conventional spatial data mining can't yield good performance. Moreover the traditional methods used for spatial data mining cannot work with MapReduce programming and unless MapReduce paradigm is used, the solution cannot be scalable and able to reap the benefits of distributed programming frameworks associated with cloud computing.

The functions of the algorithm proposed in [1] such as spatial prediction, temporal prediction, building ARIMA, obtaining hidden patterns are all time consuming with traditional spatial data mining approaches. To overcome this problem we implemented this algorithm for MapReduce, one of the distributed programming frameworks, since the spatial data mining needs more processing power. The parallel processing power bestowed by GPUs in cloud servers where Hadoop is used as MapReduce framework is optimally utilized. The framework has phases like map ad reduce. Map phase makes a set of key/value pairs and the work is given to many nodes who can work in parallel. Then the reduce phase will ensure that the results are reduced a related key/value pairs and the final result is produced. The crop yield prediction explored in [(Ceo*et al*., 2013)] is extended to meet the requirements of MapReduce programming. The implementation of the extended algorithm is made using Java programming language. The MapReduce programming when employed to the algorithm, its functionality is conceptually presented in Figure 2.

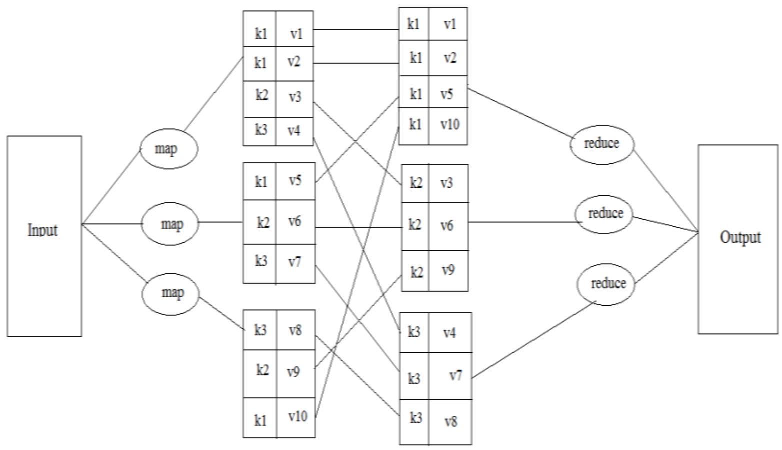

Figure 2:- MapReduce programming paradigm employed

The architecture presented in Figure 2 has provision to take big data source as input. The architecture also has provision for securing mapper in case of attacks to compromise mapper. The threat model considered is that the insider or outsider attacker tries to compromise the mapper which is part of MapReduce programming paradigm. The compromised mapper produces output that is malicious. This kind of attack is prevented as there is a component in the system that can take care of security verification. The generation of output module is responsible for actually employing our modified version of (Ceo*et al*., 2013) algorithm for producing prediction details.

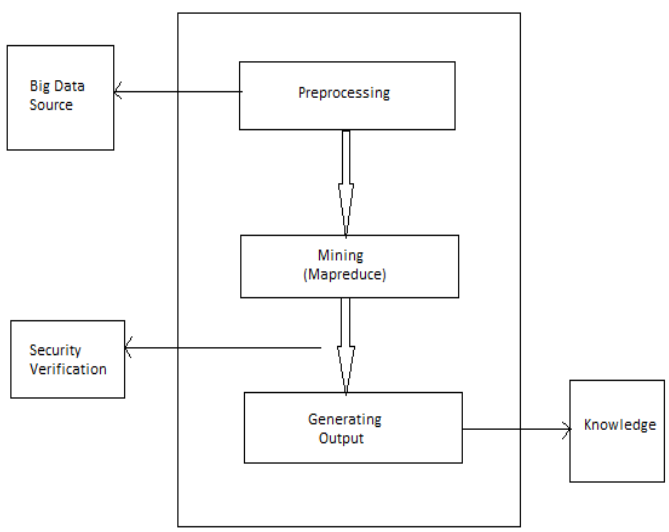

Figure 3:- Architectural overview of the proposed application

The architecture diagram represents the entire proposed application of data mining that employs MapReduce algorithm with security. In this architecture the user loads the input in the form of big data sources pertaining to crops such as Maize and Cotton besides the parameters like land fertility. Once the user loads the data then the proposed framework reads spatiotemporal data it will do the pre-processing on the data. After finishing the pre-processing the application will apply MapReduce on pre-processed data. In the middle of MapReduce process if an attacker attacks the reduce programs it will return the false data to end user. So we need to provide security for map reducing process. In the proposed application we focused on the MapReduce version of (Ceo*et al*., 2013). In addition to that we also used security mechanism for map reduce concept to provide security for input data. After finishing the mining the prototype application will generate the result and view the end-user or client. The result is nothing but crop yield prediction. As mapper is generally vulnerable to attacks launched by internal or external adversaries, that mapper part of the framework is protected in order to avoid security risks. If there is no security provided, it is possible that malicious insider or outsider can hack the system and compromise the functionality of mapper. Once the mapper is compromised, it will be under control of adversary. This will lead to output that is not consistent with inputs provided. However, the focus of this paper is on crop yield prediction in a secure fashion.

The implementation of spatial solutions towards precision agriculture is made using Spatial Hadoop [2]. Spatial Hadoop is an open source MapReduce extension designed specifically to handle huge datasets of spatial data on Apache Hadoop. SpatialHadoop is shipped with built-in spatial high level language, spatial data types, spatial indexes and efficient spatial operations. The algorithms are implemented using MapReduce programming with extended data types supported by Spatial Hadoop. The tool to demonstrate the proof of concept is implemented using Java programming language. The application and its results are explained in the ensuing section using a case study.

Outline of our Approach:-

Many researchers contributed towards crop yield prediction. The research focused on prediction of different crops. The methods used were equally different. We believed that consideration of all possible parameters in crop yield prediction can improve accuracy of prediction. This is the important hypothesis based on which we proposed the overview of approach presented in Figure 4. The parameters we considered include nitrogen, potassium, phosphorus, solar radiation, rainfall, minimum and maximum temperatures. Another distinguished feature of our approach is the usage of datasets in temporal domain. Moreover we used satellite images of the crop in order to collect live vegetation index from the image in temporal domain. The temporal difference between three datasets collected at different intervals of the crop was subjected to have vegetation index collected.

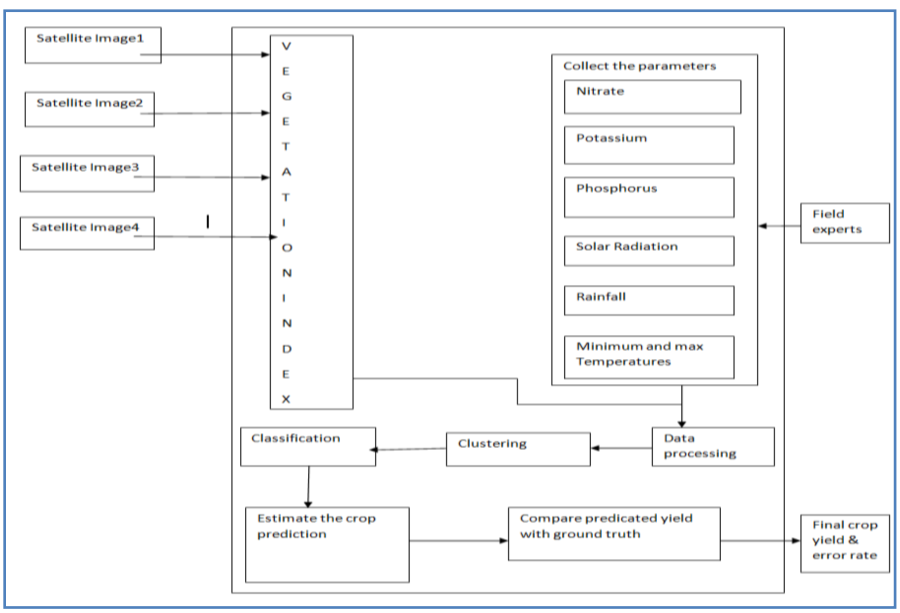

Figure 4:- Overview of the proposed approach for crop yield prediction

The datasets containing spatio-temporal data (satellite images collected at different time intervals during the crop) are subjected to the collection of different types of vegetation index. There are different indices such as Normalized Difference Vegetation Index (NDVI), Temperature Condition Index (TCI), and Vegetation Condition Index (VCI). Generally, these indices provide valuable data for determination of soil wetness, drought detection, vegetation health and evaluation of weather impacts on health and productivity of crops (Koga et al., 2003). In this paper, we considered NDVI, Soil Moisture (SM), Surface Temperature (ST), and Rain Fall (RF) besides other inputs given by field experts such as nitrate, potassium, phosphorus, and solar radiation. The annual average of NDVI, SM, ST and RF were considered in prediction process.

The overall process is as described here. From satellite datasets collected from NRSI, Hyderabad that cover cotton and maize crops at different time intervals with respect to crop seasons. The satellite images are subjected to MapReduce programming using Spatial Hadoop where the NDVI is computed in a distributed environment. The rationale behind this is that the distributed programming frameworks like Hadoop can handle huge amount of live data. Towards precision agriculture, it can help to have an integrated approach. We considered weights provided by field experts with respect to the role of Nitrogen, Potassium, Phosphorus, and solar radiation. These weights and vegetation indices are used in the prediction model.

Case study details:-

Karimnagar is a district in Telangana state of India. This district is considered as case study for crop yield prediction. Most of the people depend on agriculture in this region. The crop yield prediction in this state is the subject of this paper. This is taken as the crop yield prediction can help agriculture department to make strategic decisions. This section provides agriculture dynamics in Karimnagar and the results of crop yield prediction.

Agriculture Crops Dynamics in Telangana:-

Agriculture in Telangana state reflects many crops being cultivated. The cultivated land is distributed across the state. Food and non-food crops are cultivated. The area in which particular crop or crops are cultivated is shown in Figure 5.

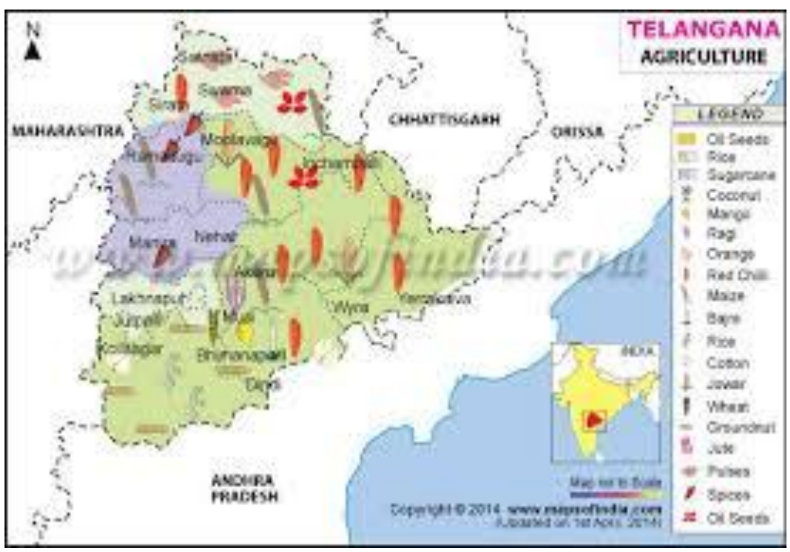

Figure 5:- Agriculture crops dynamics in Telangana

As shown in Figure 5, it is evident that the crops such as Cotton, Maize and so on are cultivated in different areas of Telangana state. However, in this paper Karimnagar area is considered for experiments. In Karimnagar Maize and Cotton crops are considered. The data is collected from officials of agriculture department of Telangana, India. The data is collected for years from 2010 to 2014.

Crops Distribution Dynamics in Karimnagar:-

The agricultural crops in Karimnagar are taken in terms of data collected from officials and the percentage of different crops cultivated in the state is computed. The distribution of crops is presented in Figure 6.

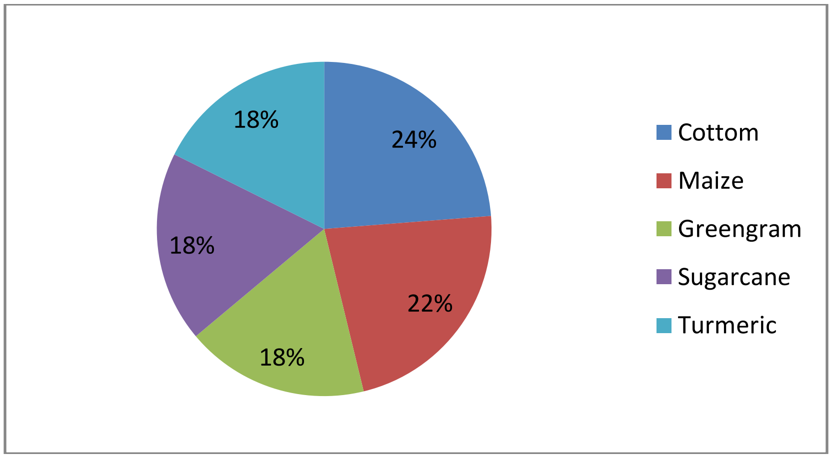

Figure 6:- Crop distribution dynamics in Karimnagar region

As shown in Figure 6, the cotton crop is cultivated more in Karimnagar. After that the second highest cultivated crop is Maize. These two are considered for experiments in this paper. The crop yield is predicted from 2010 to 2014 and the results are compared with the ground truth statistics obtained by agriculture officials.

Results:-

We took data from 2010 to 2014 of agriculture of Karimnagar region of Telangana state, India. This data considered as training sample. The parameter estimation and prediction method used are taken from (Ceo*et al.*, 2013). However, we extended the algorithm in such a way that it works for MapReduce programming framework such as Hadoop. The NDVI, ST, SM and RF values

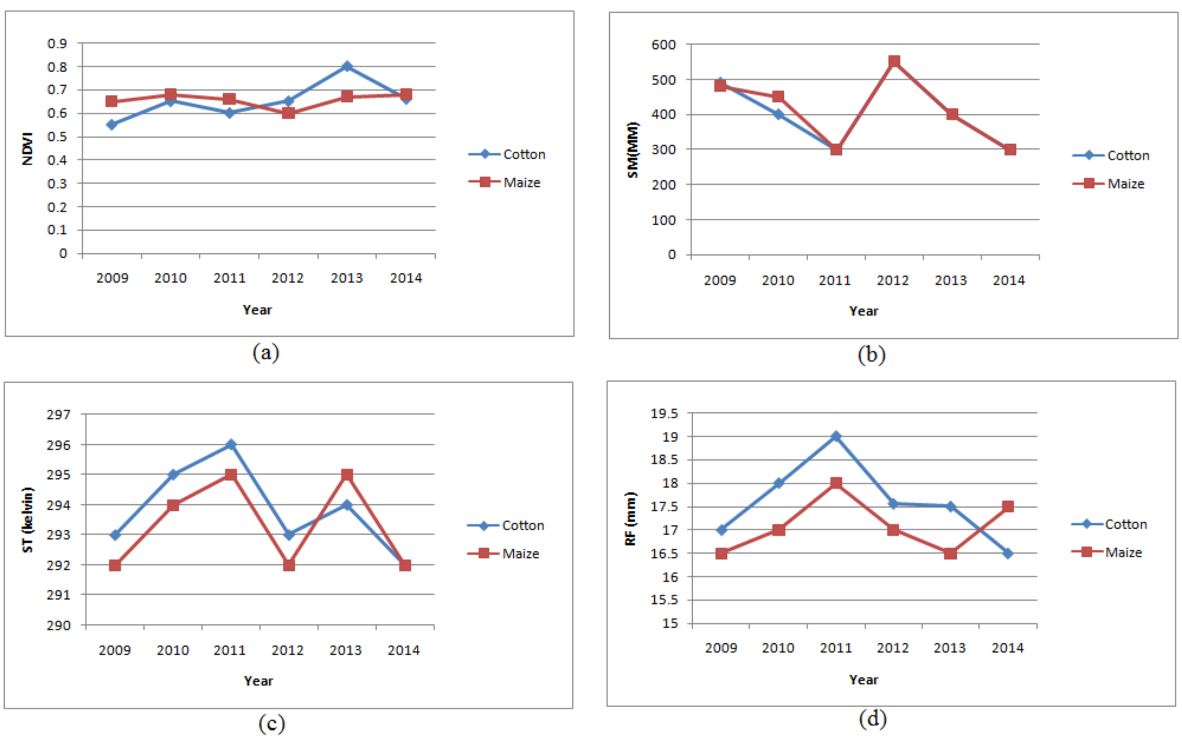

Figure 7:- NDVI, ST, SM and RF values for year 2009-2014

As shown in Figure 7, the measured values for NDVI, ST, SM, and RF were presented for 6 years. These values have their role to play in the prediction mechanism. We did our experiments with Spatial Hadoop where MapReduce programming is the underlying programming model. Taking the data from 2010 to 2014 as training sample and employing our proposed algorithm and proposed framework, the results are presented here for Cotton crop yield.

Table 1:- Prediction results for cotton crop

Year	The actual value	The predicted value	Error
2010	194355	195400	0.031
2011	194370	194320	-0.005
2012	194420	195490	0.025
2013	194460	194430	-0.007
2014	194520	195550	0.02

As shown in Table 1, the crop yield prediction results are presented for cotton crop. The experiments are made for five consecutive years from 2010 to 2014. The error rate is shown to reflect the accuracy of the method used.

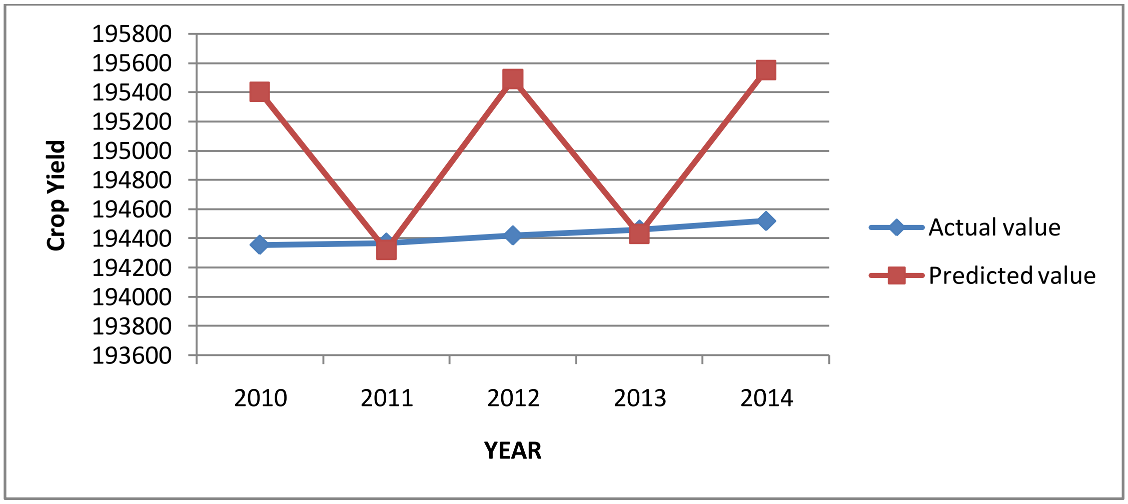

Figure 8:- Cotton crop yield prediction results

As shown in Figure 8, it is evident that the crop yield prediction results are visualized. The predicted value slightly differs from that of actual value.

Maize Yield Prediction Results:-

Table 2:- Prediction results for maize crop

Year	Actual value	Predicted value	Error
2010	183578	184600	0.025
2011	183660	184650	0.036
2012	183650	183600	-0.004
2013	183680	184720	0.041
2014	183720	183680	-0.005

As shown in Table 2, the prediction results for maize crop are shown. The error is computed and shown to reflect the accuracy of the method.

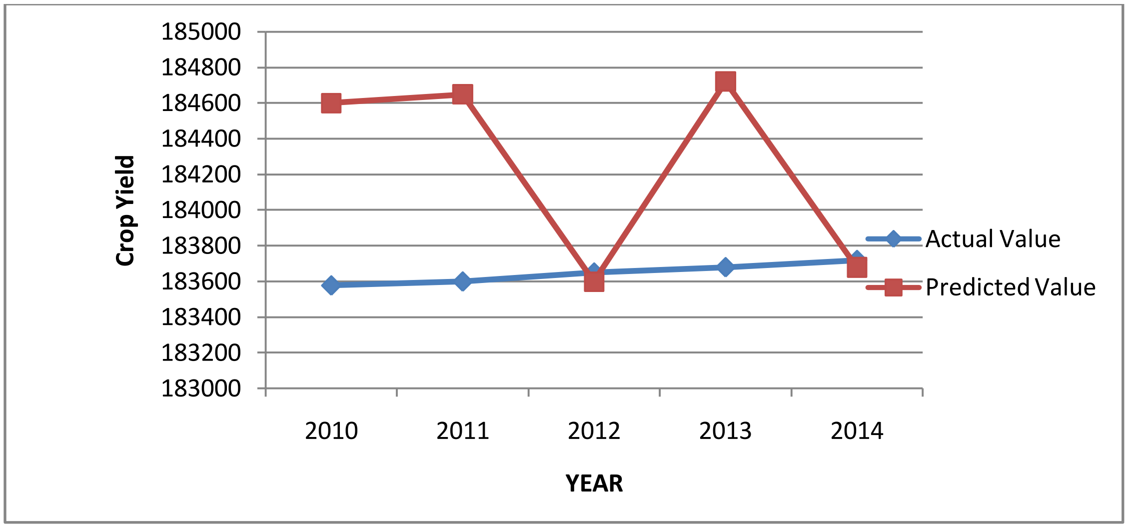

Figure 9:- Maize crop yield prediction results

As shown in Figure 9 it is evident that the maize crop yield prediction results are compared with actual results. The predicted value slightly differs from actual value reflecting the accuracy of the method.

Results after Enhancement:-

The proposed approach considers various parameters and their weights in addition to NDVI, ST, SM and RF. The enhanced approach has been tested with the same dataset. However, the initial results revealed that there is little improvement in the results. We intended to explore more on the tradeoffs on the variables. Thus it will be our future endeavour to gain more insights into the proposed method.

Table 3:- Prediction results for cotton crop.

Year	The actual value	The predicted value	Error
2010	194355	195100	0.02
2011	194370	194110	-0.07
2012	194420	194000	-0.011
2013	194460	193000	-0.039
2014	194520	195400	0.023

As shown in Table 3, the results show the predicted cotton crop yield values are compared with the actual value. The error rate is also presented which reflects the performance of the method.

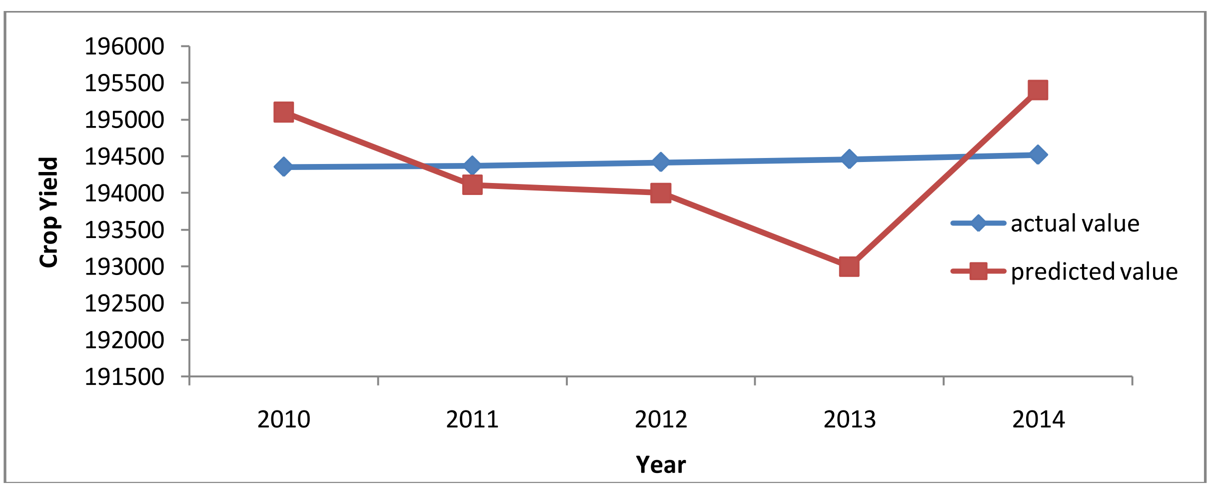

Figure 10:- Cotton crop yield prediction results with enhanced approach.

As shown in Figure 10, it is evident that the crop yield prediction for cotton crop is visualized. The results show the difference between the predicated values and actual values showing the performance of the method.

Maize Yield Prediction Results:-

Table 4:- Prediction results for maize crop

Year	The actual value	The predicted value	Error
2010	183578	184600	0.03
2011	183600	184650	0.085
2012	183650	183800	0.005
2013	183680	184720	0.031
2014	183720	183680	-0.001

As shown in Table 4 the enhanced method improved the performance of crop yield prediction for maize crop. It is evident in the reduction of error rate.

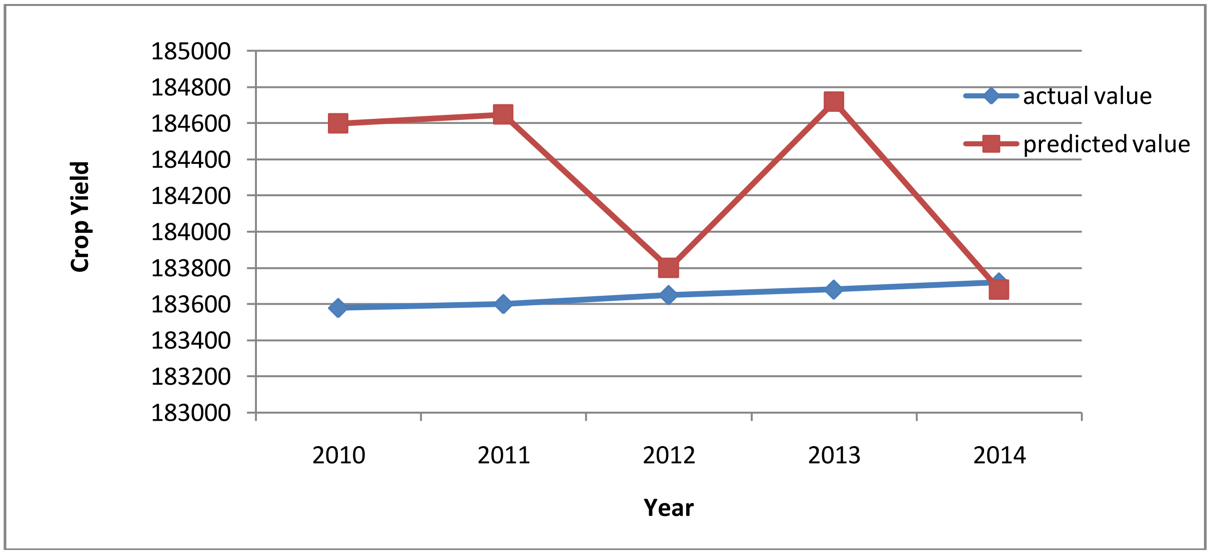

Figure 11:- Maize crop yield prediction results with enhanced approach

As can be seen in the results there is some improvement in prediction accuracy. There is possibility to reduce the prediction error further. This can be achieved by investigating the tradeoffs between the variables and the results. This is out of the scope of this paper and deferred for future work.

Discussion or Comparative Study:-

The methodology used in this paper for crop yield prediction differs from most of the approaches found in the literature. It focuses on the distributed programming framework for processing big data. Cao et al. (2013) proposed data mining approach for maize crop yield prediction. They used spatio-temporal approach. Our approach is inspired by this. However, we focused on enhancing it with other parameters besides using MapReduce programming paradigm. The datasets we used are more recent ones and they are related to two crops namely cotton and maize. The prediction method of Cao et al. was not for big data. In our work we explored satellite imagery for spatial data mining in order to obtain vital signs of crops such as vegetation indices besides using other attributes necessary for accurate prediction of crop yield. The datasets used by Cao et al. were different from ours. We used both remote sensing imagery of different periods of crops and also the ground truth datasets available in agriculture department. Our work includes both spatial and temporal data mining as the work studies different stages of crop for obtaining vegetation parameters before using them in the final prediction of crop yield.

Conclusions and Future work:-

In this paper we focused on crop yield prediction. Since the dataset contains spatiotemporal data the recent algorithm proposed in (Ceo*et al.*, 2013) is suitable for the prediction. However, we believe that spatial data mining in future needs to handle huge amount of data and such bulk of data processing and storing is not feasible in the local machine. Therefore in this paper we extended the crop yield prediction algorithm and made it compatible with

MapReduce program, a new programming paradigm, that makes use of parallel processing power of modern computing resources available through cloud computing. We proposed a framework and implemented it. Our solution not only focuses on the crop yield prediction but also throws light into security while processing the data. The dataset comprises of six years data of crops such as Cotton and Maize. The data was collected from government officials of Telangana state, India. The area considered for experiments is Karimnagar where Cotton and Maize are cultivated more. The proposed algorithm that is suitable for MapReduuce programming exploits the power of parallel processing that is exploited by distributed programming frameworks such as Hadoop. The empirical results reveal the significance of the proposed approach. The error rate is minimal. This research can be extended further to investigate the relationships and tradeoffs between the variables and crop yield prediction values towards precision agriculture.

Declarations:-

Abbreviations

ANN	Artificial Neural Network
ARIMA	Auto Regressive Integrated Moving Average
GPS	Global Positioning System
GPU	Graphics Processing Unit
IBM	International Business Machines
ITC	Information Technology and Communications
NDVI	Normalized Difference Vegetation Index
NRSI	National Remote Sensing Institute
RF	Rain Fall
SM	Soil Moisture
ST	Surface Temperature
TCI	Temperature Condition Index
VCI	Vegetation Condition Index

Ethical Approval and Consent to Participate:-
Not applicable

Consent for Publication:-
The authors declare consent for publication.

Availability of Supporting Data:-
Support data is available with authors

Funding:-
There is no funding source for this research.

Authors' Contributions:-
AM involved in the main research on crop yield prediction including dataset collection, methodology and implementation. GN contributed to identify various parameters that are to be considered for better crop yield prediction besides helping in validating of methodology. All authors read and approved the final manuscript.

Acknowledgements:-
We thank Prof.Rajan, Dr. P.K Vaishali &Dr.Konda Srinivas and for their comments which helped to improve this paper.

Competing Interests:-
The authors declare that they have no competing interests.

References:-

1. ZeluJia (2009). An Expert System Based on Spatial Data Mining Used Decision Tree for Agriculture Land Grading. *IEEE*. p4-5.
2. M.R.Pavan Kumar. (2012). International Journal of Advanced Research in Computer Science and Software Engineering. *Research Paper*. 2 (5), p1-5.
3. D.Rajesh. (2011). Application of Spatial Data Mining for Agriculture.*International Journal of Computer Applications*. 15 (2), p62-66.
4. Latika Sharma, Nitu Mehta. (2012). Data Mining Techniques: A Tool For Knowledge Management System In Agriculture. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*. 1 (5), p4-5.
5. AndreeaJulea, Nicolas M´eger, Christophe Rigotti,. (2012). Efficient Spatio-temporal Mining of Satellite Image Time Series for Agricultural Monitoring. *Transactions on Machine Learning and Data Mining*. 5 (1), p1-5.
6. YETHIRAJ N G. (2012). APPLYING DATA MINING TECHNIQUES IN THE FIELD OF AGRICULTURE AND ALLIED SCIENCES. *Integrated Intelligents Research*. 1 (2), p62-66.
7. Shital Hitesh Bhojani. (2013). Geospatial Data Mining Techniques: Knowledge Discovery in Agricultural. *Research Paper*. 3 (1), p1-5.
8. Mucherinand G. Ruß. (2012). Recent Developments in Data Mining and Agriculture. *IEEE*. p1-9.
9. Georg Ruß. (2012). Spatial Data Mining in Precision Agriculture Application Lecture @UFZ. *IEEE*. p4-5.
10. HAN-WEN HSIAO,MENG-SHU TSAI, AND SHAO-CHIANGWANG. (2006). Spatial Data Mining of Colocation Patterns for Decision Support in Agriculture. *Asian Journal of Health and Information Sciences*. 1 (1), p1-5.
11. NaniPalkhivala. (2012). Spatial Data Mining. *IEEE*,p4-5.
12. K.Venkateswara Rao, A.Govardhan and K.V.Chalapati Rao. (2012). SPATIOTEMPORAL DATA MINING: ISSUES, TASKS AND APPLICATIONS.*International Journal of Computer Science & Engineering Survey*. 3 (1), p62-66.
13. Harpinder Singh, R.K. Setia, P.K. Litoria, BrijendraPateriya. (2014). Study of Spatial and Temporal Variations in Surface Air Temperature Using Spatial Data Mining. *IEEE*. 5 (4), p62-66.
14. VibhaMaduskar, Prof. yashovardhankelkar. (2012). Survey on Data Mining. *International Journal of Emerging Technology and Advanced Engineering*. 2 (2), p1-9.
15. Intel. 2013. Planning Guide Getting Started With Big Data. *Intel IT Center*. p1-24.
16. Mike Ferguson. 2012. Architecting A Big Data Platform for Analytics. *England: Intelligent Business Strategies*. p1-36.
17. Dibyendu Bhattacharya. 2013. Analytics On Big Fast Data Using Real Time Stream Data Processing Architecture. us: *EMC Proven Professional Knowledge Sharing*. p1-34.
18. LiranEinav and Jonathan Levin. 2013. The Data Revolution and Economic Analysis. *USA: Prepared for the NBER Innovation Policy and the Economy Conference.*p1-29.
19. Michael Schroeck, Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano. 2012. Analytics: The real-world use of big data. *USA: IBM Global Business Services*. p1-20.
20. SAS. 2012. Big data Lessons from the leaders. *Economist Intelligence Unit Limited*. p1-30.
21. Oracle White Paper., 2013. *Oracle: Big Data for the Enterprise*. USA: Oracle .p1-16.
22. Australian Government., 2013. Big Data Strategy – Issues Paper.*Department of Finance and Deregulation*. p1-12.
23. Leading researchers across the United States. n.d. Challenges and Opportunities with Big Data. *Leading Researchers*. p1-17.
24. Dawei Jiang, Gang Chen, Beng Chin Ooi, Kian Lee Tan, Sai Wu. 2010.*epiC: an Extensible and Scalable System for Processing Big Data*. Singapore: CS. p1-12.
25. SAS. 2012. Big data Lessons from the leaders. *Economist Intelligence Unit Limited*. p1-30.
26. Liying Cao, Xiaohui San, Yueling Zhao, Guifen Chen (2013). "The application of the spatio-temporal data mining algorithm in maize
27. yield prediction". Elsevier, p507-514.
28. Kogan, F.N., Gitelson, A., Zakarin, E., Spivak, L., Lebed, L., 2003. AVHRR-based spectral vegetation index for quantitative assessment of vegetation state and productivity: calibration and validation. Photogrammetric Eng. Remote Sens. 69 (8), 899–906.